Optimization of Coronal Mass Ejection Ensemble Forecasting Using WSA-ENLIL with Coned Model

THESIS

Jack A. Shepherd III, First Lieutenant, USAF

AFIT-ENP-13-M-31

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENP-13-M-31

Optimization of Coronal Mass Ejection Ensemble Forecasting Using WSA-ENLIL with Coned Model

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Physics

Jack A. Shepherd III, BS, MAT

First Lieutenant, USAF

March 2013

AFIT-ENP-13-M-31

Optimization of Coronal Mass Ejection Ensemble Forecasting Using WSA-ENLIL with Coned Model

Jack A. Shepherd III, BS, MAT
First Lieutenant, USAF

Approved:

Ariel O. Acebal, Lt Col, PhD, USAF (Chairman)	5 Mar 13 Date
Matthew B. Garvin, Capt, PhD, USAF (Member)	4 Mar 13 Date
Antti A. Pulkkinen, PhD, NASA/GSFC (Member)	02-25-13 Date
Aleksandre Taktakishvili, PhD, NASA/GSFC (Member)	02/25/13 Date

Abstract

The combination of the Wang-Sheeley-Arge (WSA) coronal model, ENLIL version 2.7, and the Coned model version 1.4 was utilized to form ensemble forecast for 21 coronal mass ejections (CMEs). The input parameters for WSA-ENLIL were taken from 100 sets of CME measurements automatically derived from the Coned model using LASCO C3 difference images along with a bootstrap technique. The Coned model was improved by adding a weight for the associated flare location to push the propagation axis towards the flare location and an additional image was used. The CME propagation time forecasts utilizing the improved Coned model outperformed previous versions by a large margin. The mean absolute forecast error of the median ensemble results was improved by over 43% over the original Coned model version 1.3, placing the arrival time within 4.59 hours. The arrival time forecasts for 12 of the 21 events fell within the ensemble average plus or minus one standard deviation and 19 of the 21 events had the actual propagation time within the range of the ensemble. The model was also used to look at the propagation of multiple CMEs within a 48-hour period. This resulted in an improvement in the error of a sample CME from 8.09 hours to 1.77 hours.

I would like to dedicate this paper to my grandfather who died before he could see the results of this research. His love for space helped to spark my own interests and without him, I would have never followed this path I am now on.

Acknowledgments

First, I would like to thank my wife for still being there when I would come back from days of working on my research. Thank you for making me food when I forgot to eat, reminding me to sleep when I forgot to sleep, and forcing me to go outside and see the Sun on occasion. Without you, I don't know how I could have stayed sane.

I would also like to thank Lt Col Acebal for being one of the best advisors that I could imagine. Thanks for pointing me in the right direction when I needed it and providing the background material I needed to muddle my way through all the problems that appeared throughout my research.

Finally, I would like to thank Antti and Sandro at CCMC for giving me the opportunity to work on such an incredible project. Without Sandro, I could have never gotten ENLIL to work correctly and Antti helped explain all the intricacies of the Coned model the way that no one else could. It has been amazing working with you guys and I appreciate all that you have done for me.

Jack A. Shepherd III

Table of Contents

Page

List of Figures

List of Tables

OPTIMIZATION OF CORONAL MASS EJECTION ENSEMBLE FORECASTING USING WSA-ENLIL WITH CONED MODEL

I. Introduction

Coronal Mass Ejections (CMEs) are the largest explosions in the solar system, ejecting up to 10^{13} kg of mass with velocities of 1000 km/s or more [*Chen*, 2011]. These CMEs have a chance to impact Earth and result in hazardous space weather conditions that can have disruptive effects on communication [*Tascione*, 1994], our space assets [*Afraimovich et al.*, 2003], and electrical systems on the Earth's surface [Boteler et al., 1998]. These reasons make the prediction of arrival times and impacts of CMEs on Earth of great interest to the United States Air Force and NASA. This research will focus on improving the forecasting accuracy of the arrival times of these CME effects on Earth.

In order to predict the arrival time of the CME, the Wang-Sheely-Arge (WSA)-ENLIL model with the Coned model will be utilized. ENLIL is a time-dependent three-dimensional magnetohydrodynamic (MHD) model that simulates the global behavior of a plasma [Odstrcil, 2004]. WSA calculates the characteristics of the solar wind in the heliosphere from magnetogram measurements and is used for the inner boundary conditions for ENLIL. The Coned model [Pulkkinen et al., 2010] uses a bootstrap approach and pattern recognition technique to capture CMEs in LASCO C3 difference images and fits a cone approximation to the images in order to characterize a CME. This characterization is then used as an input parameter for ENLIL in order to predict the evolution of the CME and background solar wind from the Sun to the Earth.

In order to produce an accurate representation of the estimated time of arrival and impact, an ensemble of runs is performed with a spread imposed upon the input parameters. The most recent version of the Coned model is version 1.3 and it tends to push all CMEs so they are directed towards the Earth [Emmons, 2012]. Version 1.3 of the Coned model has a mean absolute propagation time forecast error of 9.06 hours [Emmons, 2012]. In order to reduce this forecast error, this research inquires whether adding a weight to propagate the CME radially outward from the flare location and adding an additional image will produce improved forecasting of the CME impact on Earth. These changes are part of version 1.4 of the Coned model.

This analysis applied an ensemble forecasting technique to 15 CMEs using the WSA-ENLIL with Coned model. The ensembles were created using 100 sets of initial states that were derived from the Coned model version 1.4 which were then used as inputs to WSA-ENLIL version 2.7 to create distributions of predicted propagation times for the CMEs. The 15 CMEs used by Emmons [2012] were used in order to determine the improvements that were made to the model. Then, six additional CMEs were analyzed in order to verify the improvements.

This analysis was then repeated by removing a climatological weight for the opening half angle of the CME cone. This tested the robustness of the model by using only information gained from the actual CME rather than historical averages of previous CMEs. Additionally, a single case was analyzed utilizing two CMEs at the same time in order to test the capability to perform the calculation as well as any improvements that could be made by including the additional CME in the calculations.

The remainder of this document is structured such that chapter 2 provides the background for this analysis and includes a discussion on CMEs, the WSA-ENLIL with Coned model, and a look into previous research using WSA-ENLIL with Coned model. Chapter 3 provides the methodology that was used in the analysis. Chapter 4 contains the statistical analysis and discussion of the results. Finally, Chapter 5 discusses the conclusion of the analysis.

II. Background

Chapter Overview

The purpose of this chapter is to provide background of the models and techniques used to generate ensemble coronal mass ejection forecasts. The first part of this chapter describes coronal mass ejections and their propagation through the interplanetary magnetic field. Next, the Wang-Sheely-Arge, ENLIL, and Coned models are described. Finally, previous works on ensemble coronal mass ejection forecasts are highlighted.

Coronal Mass Ejections

Overview

Coronal mass ejections (CMEs) are energetic events that occur on the surface of the Sun which eject enormous amounts of plasma and their associated magnetic fields into interplanetary space. These explosions on the surface of the sun are the largest eruptions in our solar system. They are relatively common but occur much less during solar minimum than during the solar maximum. A CME associated with one of the largest solar flares ever recorded is shown in Figure 1. Gopalswamy et al., [2003] found that during the solar minimum, a CME happens, on average, every other day. The average rate which a CME occurs slowly increases until it reaches a peak of about 6 CMEs per day during the solar maximum.

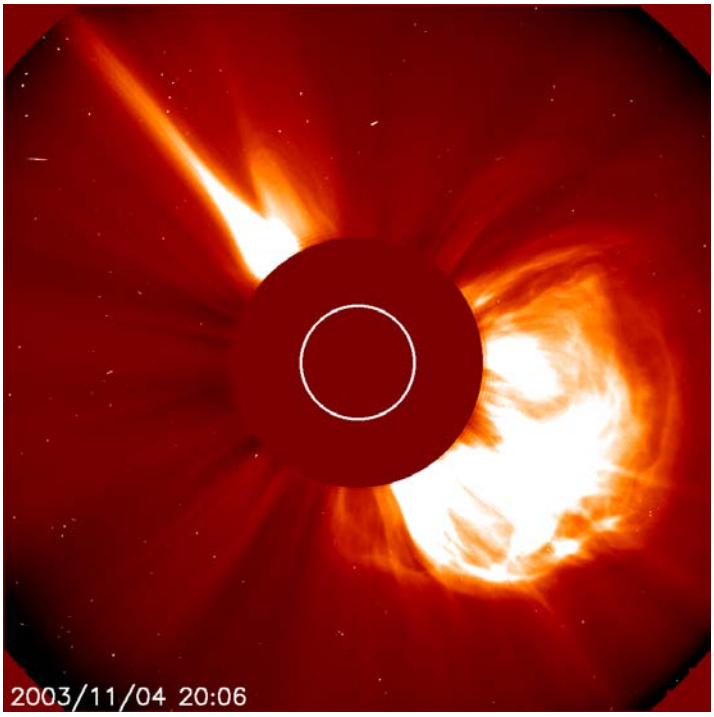

Figure 1: Image of a CME eruption associated with the largest modern solar flare recorded (~ X28) that occurred 4 November 2003 from LASCO C2. (NASA)

Eruption

The explosion associated with a CME releases massive amounts of energy. According to Emslie et al. [2004], the total kinetic and potential energy released in a CME is typically between 10^{22} and 10^{25} J. For comparison, the BP Statistical Review of World Energy [2012] estimates that the total worldwide energy consumption was greater than 5 x 10^{20} J during the calendar year 2011. This means that the energy released in a single CME ranges from 20 to 20,000 times the world's yearly energy consumption.

CMEs, much like solar flares, involve the conversion of one type of energy to another. Chen [2011] estimated the energy density of CMEs to range from 0.01 to 10 J/m^3 by estimating the typical CME volume of 10^{24} m^3. Table 1 shows estimates of typical energy densities observed for different energy sources in the solar corona. The

magnetic energy density is far greater than any of the other sources. Since the typical energy densities of CMEs reach as high as 10 J/m^3, the primary source of energy of these CMEs must be magnetic. None of the other energy densities reach values high enough to produce the observed values.

Table 1: A list of the estimated coronal energy sources adapted from Forbes [2000].

Form of Energy	Energy Density (J/m^3)	Observed Averaged Value
Kinetic ($\frac{1}{2} m_p n V^2$)	8×10^{-4}	$n = 10^{15}$ m^{-3}, $V = 1$ km/s
Thermal (nkT)	1×10^{-2}	$T = 10^6$ K
Gravitational ($nm_p gh$)	5×10^{-2}	$h = 10^5$ km
Magnetic ($B^2/2\mu_0$)	40	$B = 10^{-2}$ T

The angle which a CME is projected in the plane of the sky spans nearly the full range of possible values. For example, Yashiro et al. [2004] found CMEs exhibiting an angular width from as low as 2° to as high as 360°. The distribution of these angular widths is used to generate the dividing lines between the narrow CMEs and the normal CMEs although the exact numerical values of these dividing lines are not agreed upon. Chen [2011] uses 10° as the dividing line between narrow and normal while Yashiro et al. [2004] use 20° to 120° to define normal and call any CME below that range narrow and any CME above wide. Regardless, around 75% of CMEs fall within the 20° to 120° range with slightly more falling below that range than above it. This is illustrated in Table 2 which shows the total number of CMEs that occurred for each year from 1996 to 2002 and the percentage of each CME that fell within the narrow, normal, or wide range using the 20° and 120° upper and lower bounds, respectively, for the normal type of CME.

Table 2: List of the number of narrow, normal, and wide CMEs adapted from Yashiro, et al. [2004].

Year	Total CMEs	Narrow	Normal	Wide
1996	204	16%	77%	6%
1997	351	11%	79%	9%
1998	697	20%	70%	9%
1999	957	15%	71%	13%
2000	1580	21%	68%	10%
2001	1465	13%	73%	14%
2002	1652	23%	67%	10%

Normal CMEs exhibit a closed loop pattern. This is due to the motion of the plasma along magnetic field lines, but in this case, the closed field lines create a closed loop structure which can be seen in Figure 2 (a). Normal CMEs typically have a three part structure composed of a bright core on the inside, followed by a dark cavity, and surrounded by a bright loop [Illing and Hundhausen, 1985]. Despite this three part structure representing the standard morphology for CMEs, Webb and Hundhausen [1987] found that only about 30% of CMEs have all three parts of the structure.

The narrow CME generally exhibits a jet-like effect where material is ejected from the sun in a very narrow stream. This streaming effect is likely due to the plasma travelling along open magnetic field lines along the solar surface where the instability occurred [Chen, 2011]. This narrow type of CME is believed to be formed from the magnetic reconnection between small magnetic dipoles [Wang et al., 1998]. An example of this type of CME can be seen in Figure 2 (b) where the very small angular width is apparent.

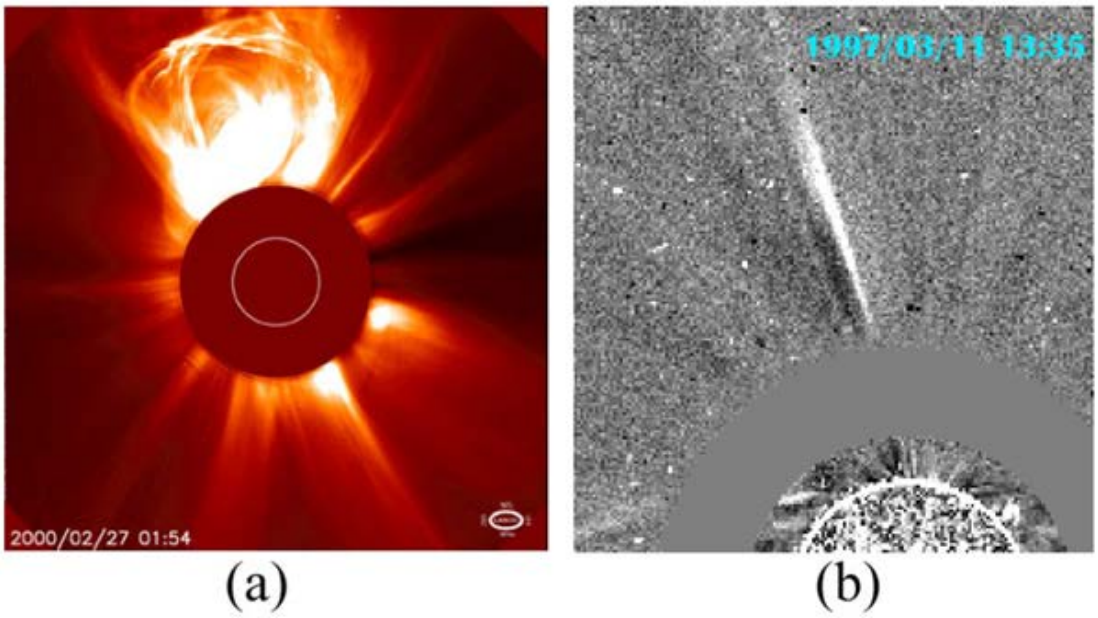

Figure 2: Picture showing the difference between the (a) normal type of CME from a LASCO C2 image and the (b) narrow type of CME from a difference image adapted from Chen [2011].

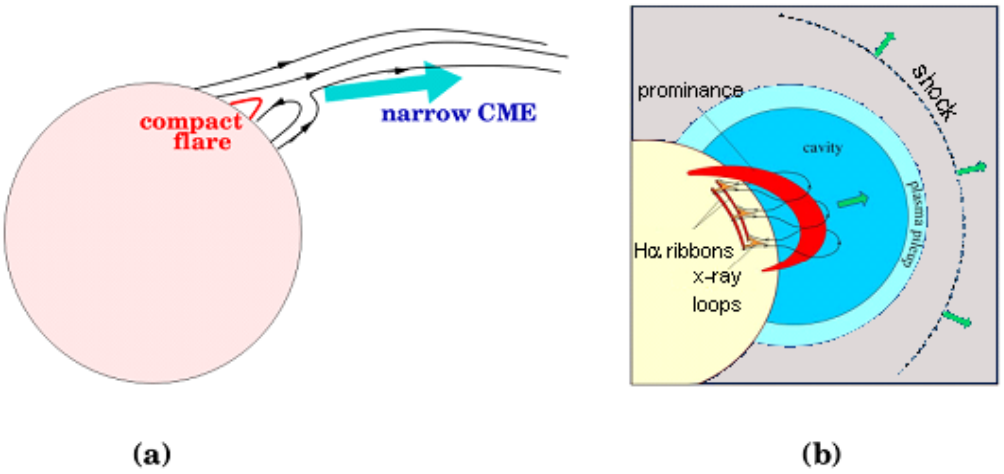

Figure 3: Picture illustrating eruptions of the (a) narrow type of CME from Chen [2011] and the (b) normal type of CME adapted from Forbes [2000].

CMEs are generally associated with solar flares although the links between them are not causal. There are rare cases where CMEs do not have a visible flare associated with them. This may be due to the flare being behind the solar limb or that the associated flare below the CME is so weak that it was not registered as a flare [Zhou et al., 2003]. While nearly all CMEs appear to be associated with solar flares; many solar flares are not associated with CMEs. This occurs more often for weaker solar flare events than for the more powerful solar flares. Wang and Zhang [2007] found that while ~90% of X-class flares are associated with a CME, the weaker M-class flares only produce a CME ~56% of the time and the even weaker C-class flares produce a CME ~30% of the time.

Another feature of CMEs is the amount of material that CME released. CMEs vary widely in mass but have an average of about 3×10^{12} kg and generally fall between 1×10^{11} and 4×10^{13} kg [Jackson, 1985]. About 15% of CMEs were found to fall below that range, however, while less than 1% was found to fall above this range [Vourlidas et al., 2002]. This mass range was studied for 2449 CMEs from 1996 to 2000 with the mass distribution shown in Figure 4. This mass is generally estimated using the Thomson-scattering formula [Chen, 2011].

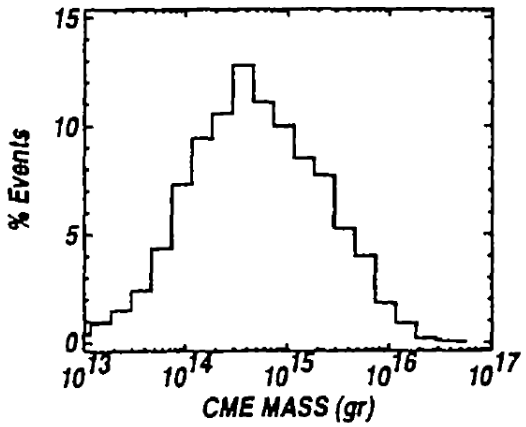

Figure 4: Histogram showing the distribution of mass of CMEs between 1 January 1996 and 30 July 2000 adapted from Vourlidas et al. [2002].

Propagation

The CME velocity profile can be divided into three phases which Zhang [2001] calls the initiation phase, the impulsive acceleration phase, and the propagation phase. The first phase is the initiation phase and begins when the CME front is first formed and undergoes a slow expansion. The speed of this expansion was found to vary from 5 to 80 km/s and lasts from 0.5 to 2 hours. If there is an associated flare this phase will occur before the onset of any associated flare and is differentiated from the second phase by having an acceleration rate that is two orders of magnitude smaller [Zhang et al., 2001]. The impulsive acceleration phase is next and occurs almost simultaneously with the flare's rise phase, if there is an associated flare. This phase usually lasts for a few to tens of minutes. The CME in this phase rapidly accelerates as fast as 3,270 m/s^2 [St. Cyr et al., 1999]. The propagation phase is the final phase and is characterized by a nearly constant velocity. It occurs after the main acceleration of the CME has concluded, near

the peak time of the soft X-ray flares, and the CME shows only relatively small increases or decreases in speed in this phase [Zhang et al., 2001].

Once the CME enters the interplanetary medium, it is often referred to as an interplanetary coronal mass ejection (ICME). The CME propagates into and through the interplanetary medium and is then assimilated into merged interaction regions in the outer heliosphere where it loses its identity [Forbes et al., 2006]. There are two general approaches used to describe the CME propagation phase while it is in the interplanetary medium. The first approach is an analytical formulation that specifies the equations that describe the motion of the CME that is undergoing acceleration and deformation forces through the solar wind. The position of the CME and its geometry are determined as a function of time through ordinary differential equations. The second method uses magnetohydrodynamic (MHD) simulations of the CME and its surroundings and uses partial differential equations that specify the motion field and the force fields at every point of a simulation grid, instead of the center of mass of the CME [Forbes et al., 2006].

Once the CME is ejected, there is a large spread in possible initial velocities as the CME heads into the interplanetary medium. These velocities can be as low 20 km/s or as high as 3500 km/s [Yashiro et al., 2004]. CMEs with initial velocities greater than the solar wind will decelerate while propagating. CMEs with initial velocities less than that of the solar wind will accelerate [Gopalswamy et al., 2000]. An empirical formula for the amount of acceleration, a, of the CME in route to 1 AU was determined to be

$$a[\mathrm{m/s^2}] = 1.41 - 0.0035\ v[\mathrm{km/s}] \qquad (1)$$

where deceleration is represented by a negative value for *a*, and *v* is the plane-of-the-sky CME speed measured [Gopalswamy et al., 2000]. For example, if the measured velocity of the CME was 200 km/s, which is slower than the background solar wind velocity, the expected acceleration would be 0.71 m/s^2. Conversely, if the measured velocity of the CME was 1500 km/s, much faster than the background solar wind velocity, the expected acceleration would be -3.84 m/s^2.

While the CME propagates outward, it also expands. An empirical relation was found by Owens et al. [2005] such that

$$V_{\mathrm{EXP}}[\mathrm{km}/s] = 0.266V[\mathrm{km}/s] - 70.61 \qquad (2)$$

where V_{EXP} is the rate CME radius is expanding and V is the velocity of the leading edge of the CME. By the time the CME reaches 1 AU, the radial dimension of the CME is typically between 0.20 and 0.25 AU [Klein and Burlaga, 1982].

Impact

The magnetic field of the CME can have a profound effect on the Earth's magnetosphere and produce severe geomagnetic storms depending on the direction and magnitude of the magnetic field associated with the CME as well as the duration of the CME impact with the Earth. If the CME has a magnetic field pointing southward with respect to Earth, then the impact will be the largest. These geomagnetic storms usually last for one to three days and have energy dissipation rates several times greater than the usual energy transfer rate from the solar wind to the magnetosphere; as high as 10^{12} Watts [Prölls, 2004].

One of the methods used to determine the impact that a CME had on the Earth's magnetosphere is the K-index. The K-index (the K comes from the German "Kennziffer" which means "index") is a ground-based measure of the magnetic activity at mid-latitudes caused by the solar wind [Tascione, 2010]. It is used as a quasi-logarithmic measure of the variation of the magnetic field from a standard measurement of the magnetic field of the Earth in calm conditions using numbers from 0 to 9 [Tascione, 2010]. A K index of one represents calm magnetic conditions while a K index of five or higher indicates a geomagnetic storm. The K-index stops at nine and that represents the most severe geomagnetic storms.

The most commonly used version of the K-index is the Kp index (the "p" stands for "planetary" so Kp literally means "planetary index") which is generated with a time resolution of three hours [Prölls, 2004]. The Kp index is calculated from the combination of K-index measurements made at 13 different location worldwide between geomagnetic latitudes of 48° to 63° [Tascione, 2010]. The Kp index indicates deviations of the Earth's magnetic field which may be caused by enhancements to space currents.

Onboard the Advanced Composition Explorer (ACE) is the Solar Wind Electron Proton Alpha Monitor (SWEPAM) and a magnetometer (MAG) instrument. SWEPAM measures the solar wind plasma electron and ion fluxes while the MAG instrument gives the IMF direction and magnitude. ACE data is used to determine the arrival time of a CME on Earth as well as estimating the impact the CME will have on the Earth.

Measurement

The photosphere of the Sun is so bright that viewing corona is difficult. The corona and CMEs are very tenuous compared to the photosphere and chromosphere of

the sun and so, in order to view these more tenuous structures, the much brighter parts of the Sun must be blocked out. A coronagraph is used in order to accomplish this goal. A coronagraph is an observational device that uses an occulting disc in order to block the direct light from the sun allowing the surrounding structures to be seen more clearly. The first optically observed CME was recorded by a coronagraph aboard NASA's Orbiting Solar Observatory 7 in December 1971 [Rycroft and Runcorn, 1973].

There is currently a coronagraph aboard the Solar and Heliospherical Observatory (SOHO); a joint project between the European Space Agency (ESA) and NASA launched December 1995. One of the payloads aboard SOHO is the Large Angle and Spectrometric Coronagraph (LASCO) that takes visible spectrum images of the solar corona between 1.1 and 32 solar radii. LASCO is made up of three different telescopes, C1, C2, and C3, which each observe a different location around the sun. The C1 telescope observed within the 1.1 to 3 solar radii range but no longer works. The C2 telescope can image between 1.5 and 6 solar radii. The C3 telescope has the largest range from 3 to 32 solar radii. Difference images can also be used to remove constant background features in order to make events, such as a CME, more easily locatable. These difference images can be used to determine the location of a CME front as well as to estimate its velocity. One of the LASCO C3 difference images used in this study is shown in Figure 5.

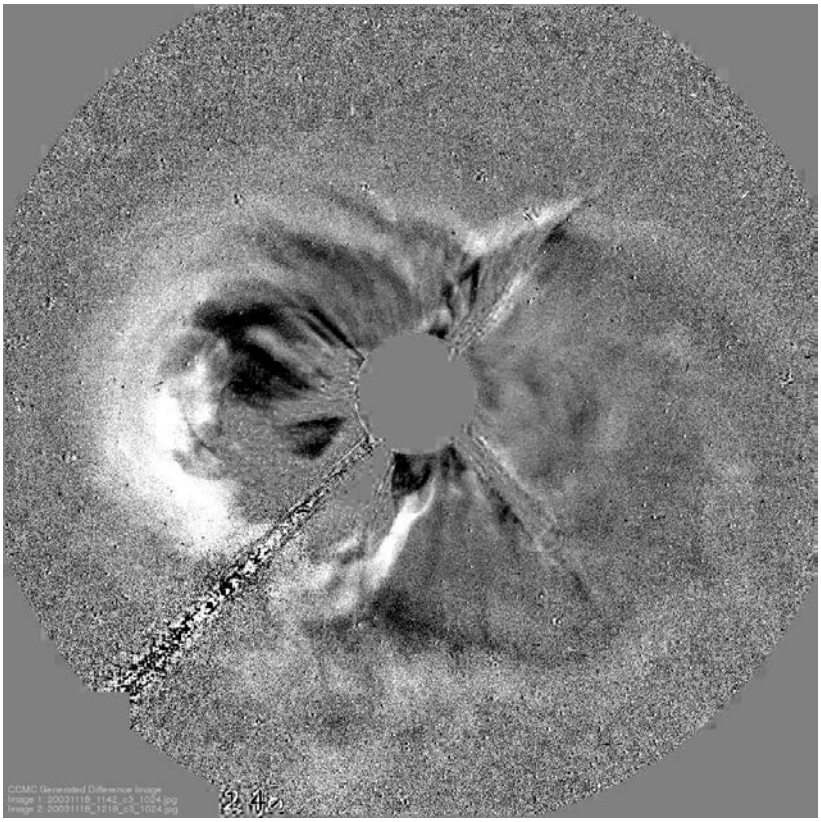

Figure 5: A LASCO C3 difference image of the 18 November 2003 CME.

Wang-Sheeley-Arge and ENLIL with Coned Model

Wang-Sheeley-Arge (WSA) and ENLIL with Coned model solves the MHD equations in order to predict how a CME will behave after leaving the sun. The Coned model automatically calculates CME characteristics (velocity, opening angle, and propagation axis) as inputs for ENLIL while the WSA coronal model provides the boundary conditions and the magnetic characteristics of the solar wind. ENLIL is then run to calculate the propagation of the CME given the inputs and to determine if the CME will impact the Earth and, if so, what effects it will have. Here, the Coned model, WSA, and ENLIL are examined in order to gain an understanding of how the linked models work together.

Coned Model

The Coned model, sometimes called the Automatic Cone model, is used to produce parameters of the CME automatically which can then be input into ENLIL. The Cone model is a mathematical construct where a CME is assumed to have the shape of a cone in order to simplify the calculation of these parameters which are the CME velocity, propagation axis, and opening angle. The Coned model was created in 2009 by Pulkkinen et al. and it automatically calculates the cone parameters from the Cone model.

The parameters of the CME cone that are of interest are shown in Figure 6. The plane (y', z') defines the plane of sky and is perpendicular to the x' axis which points towards the Earth. The angle α defines the direction of propagation of the CME in the (y', z') plane which is the angle between the y' axis and the x axis as projected into the (y', z') plane. The angle θ defines the rotation of the CME cone off of the (y', z') plane which also represents the angle between the x' axis and the x axis. The angle ω defines the opening half-angle of the CME cone, x_0 is the initial distance of the CME cone front in the rotated coordinates (x, y, z), v is the velocity of the propagation of the CME cone front, and Δt is the time interval during which the cone front propagates from x_0 to x.

The Coned model uses a time series of LASCO C3 difference images and then utilizes a three step image processing algorithm on the images to automatically determine the location of the CME front in each of the images. The first step is to add contrast to the image by linearly mapping the original values to values covering the full grayscale intensity range. Second, the image is filtered and a 25 x 25 neighborhood is used to

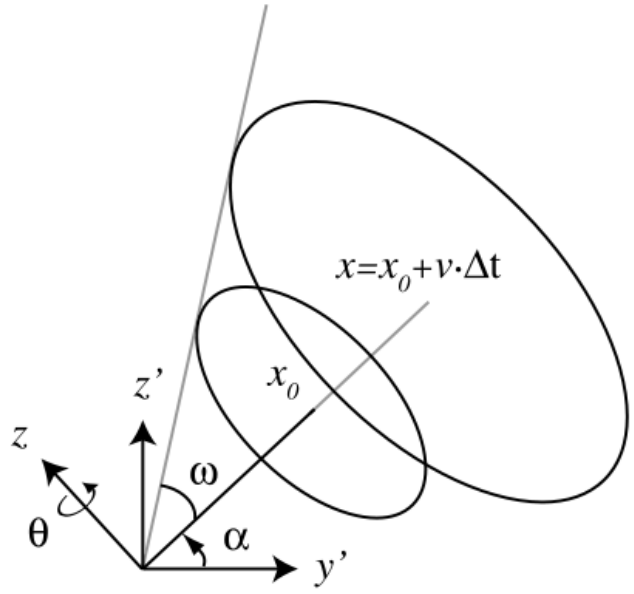

Figure 6: A representation of the parameters of the CME cone that required for the Coned model, adapted from Pulkkinen et al. [2010].

compute a median value that is then assigned to individual pixels. Finally, the pixels of the filtered image are converted into binary values based on a brightness threshold defined by the user as a certain percentage of the maximum intensity [Pulkkinen et al., 2010]. Pixels that are brighter than this percentage of the maximum intensity are defined as part of the CME mass are have the pixel turned on so that it is shown in white. Pixels dimmer than this percentage of the maximum intensity are determined to be part of the background and have the pixel turned off so that it is shown in black. An example of this process is shown in Figure 7.

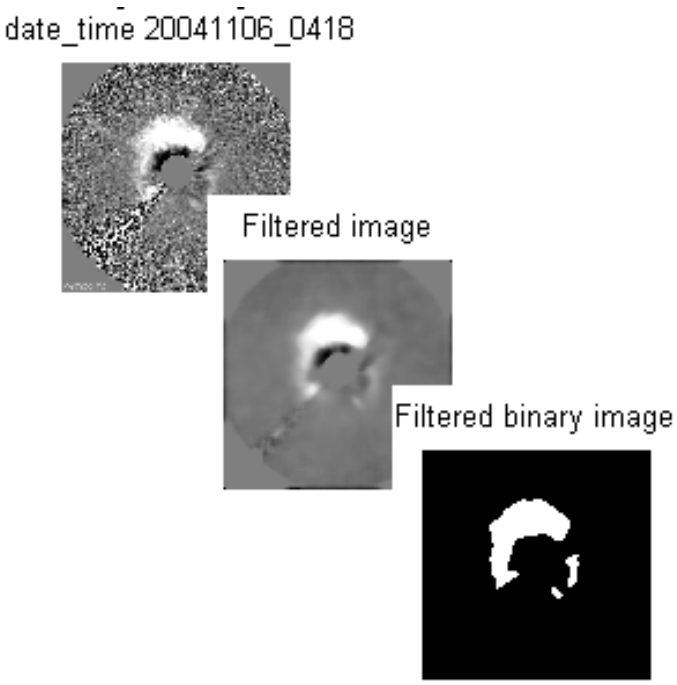

Figure 7: Example of the Coned model image processing technique utilized on a CME from 6 November 2004 utilizing LASCO C3 difference images.

After the location of the CME masses have been determined from the LASCO C3 difference images, the cone model parameters are determined from the data. First, the center of mass of all the data is computed by

$$y'_m = \frac{1}{N}\sum_i^N y'_i, \tag{3}$$

$$z'_i = \frac{1}{N}\sum_i^N z'_i, \tag{4}$$

where the summation is over all *N* data points of the CME mass [Pulkkinen et al., 2010]. Next, the direction of the propagation, α, of the CME in the (*y'*, *z'*) plane can be calculated by

$$\alpha = \tan^{-1}(z'_m / y'_m). \tag{5}$$

The data are then rotated by an angle, $-\alpha$, about the x' axis. In order to compute the remaining four parameters, $\{\theta, \omega, x_0, v\}$, an inversion scheme is invoked that is expressed as

$$\min_{\{\theta,\omega,x_0,v\}} \left[\sum_i^N \sqrt{(\hat{y}'_i - y'_i)^2 + (\hat{z}'_i - z'_i)^2} + \mu|\omega - \omega_0| \right], \tag{6}$$

where $(\hat{y}'_i, \hat{z}'_i)$ are the coordinates of the CME cone front, (y'_i, z'_i) are the coordinates of the CME mass data, μ is a weighting for the measurement of the climatological opening half-angle and was set to 3 x 10^9, and ω_0 is a climatological opening half-angle [Pulkkinen et al., 2010]. The coordinates $(\hat{y}'_i, \hat{z}'_i)$ from Equation (6) are computed by

$$\begin{aligned} & R_z^T(\theta) \cdot \begin{pmatrix} x \\ x\tan(\omega)\cos(\gamma) \\ x\tan(\omega)\sin(\gamma) \end{pmatrix} \\ &= \begin{pmatrix} x\cos(\theta) - x\tan(\omega)\cos(\gamma)\sin(\theta) \\ x\sin(\theta) + x\tan(\omega)\cos(\gamma)\cos(\theta) \\ x\tan(\omega)\sin(\gamma) \end{pmatrix} \\ &= \begin{pmatrix} \hat{x}'(\gamma) \\ \hat{y}'(\gamma) \\ \hat{z}'(\gamma) \end{pmatrix} \end{aligned} \tag{7}$$

where the operator $R_z^T(\theta)$ rotates the parameterized (as a function of angle γ) representation of the cone by the angle θ about the z axis, $x = x_0 + v\Delta t$, v is the velocity of the cone front propagation, and Δt is the time that it takes the CME cone front to propagate from x_0 to x [Pulkkinen et al., 2010]. For simplicity, it is assumed that the CME front propagates with a constant velocity between images and $(\hat{y}'_i, \hat{z}'_i)$ in Equation (6) are determined from $(\hat{y}'_i(\gamma), \hat{z}'_i(\gamma))$ from Equation (7) by selecting the angle γ that minimizes the distances to the data point (y'_i, z'_i) [Pulkkinen et al., 2010].

Equation (6) is solved using the MATLAB Optimization Toolbox version 4.1. A stabilizing factor was found to be necessary in order to prevent the results from varying by too much from run to run and so the additional term $\mu|\omega-\omega_0|$ was added in Equation (6) [Pulkkinen et al., 2010]. The climatological value ω_0 was chosen as 30° based on the CME statistics analyzed by Cyr et al. [2000] and Yashiro et al. [2004]. Finally, our values for α and θ are used to calculate the heliocentric coordinates by

$$\lambda = \frac{\pi}{2} - \cos^{-1}\big(\sin(\theta)\sin(\alpha)\big), \qquad (8)$$
$$\phi = \tan^{-1}(\tan(\theta)\cos(\alpha)),$$

where λ is the heliocentric latitude and ϕ is the heliocentric longitude.

A bootstrap method is utilized in order to determine the confidence intervals for the calculated cone parameters. This bootstrap method randomly draws subsets of data from the original set of detected CME masses and calculates the model parameters for each subset. An example of the output to this bootstrap method is shown in Figure 8 where the cone parameters are calculated from the filtered binary image. First, each time series image for a CME is processed into a filtered binary image as in Figure 7. Then 300 points are randomly selected from each image. The progression of these randomly selected points over the time series of images represents the CME propagating outwards from the Sun. This change of the CME over the time series examined is used to calculate the velocity, opening angle, and propagation axis of the CME by minimizing Equation (6). This analysis is then repeated 100 times to create a distribution of the cone parameters that can be used as input parameters for ENLIL for ensemble forecasting of the propagation time of a CME to Earth and the impact that the CME would have on the Earth's magnetosphere, which was not performed in this study.

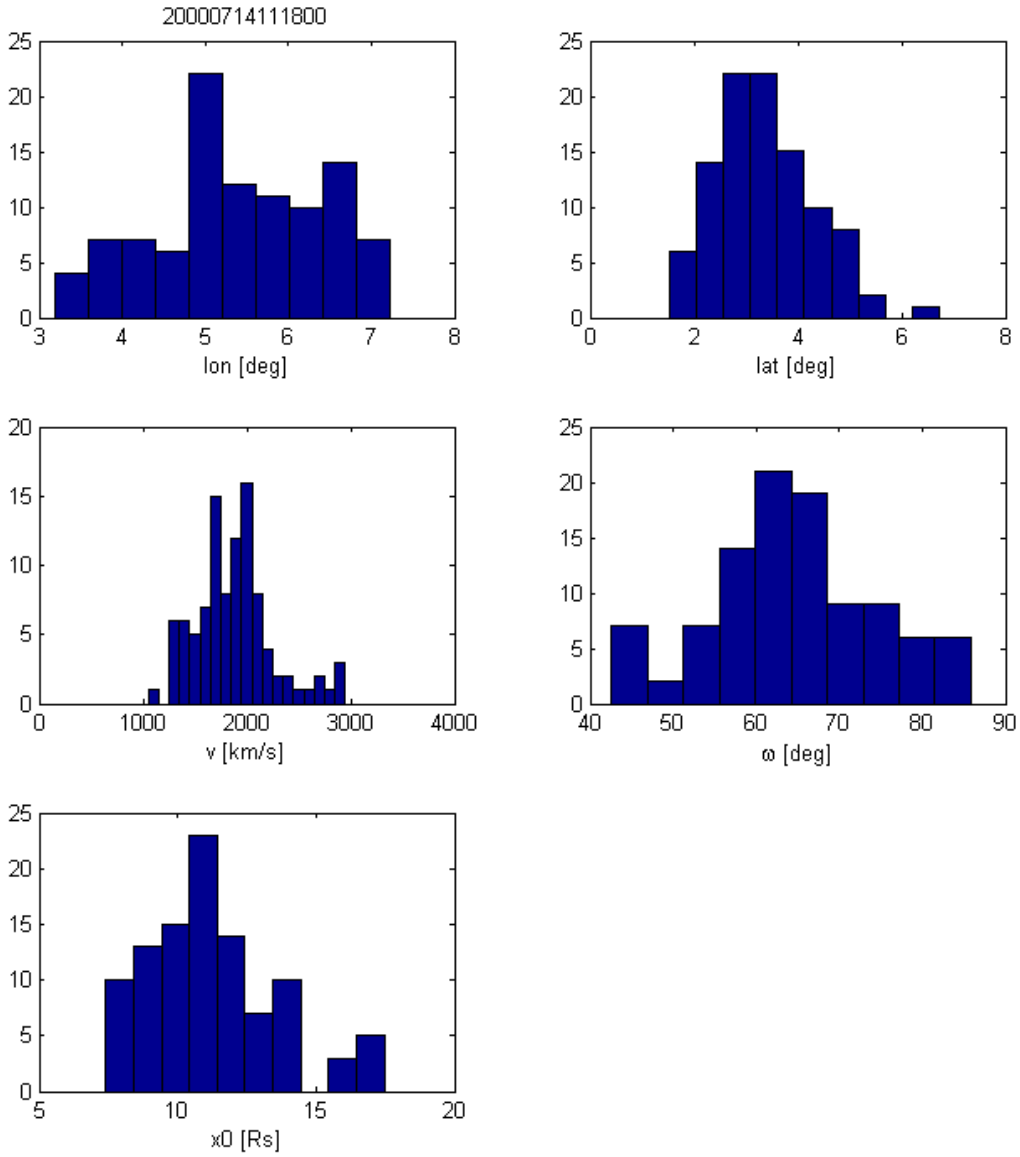

Figure 8: Distribution of the cone parameters obtained from using the bootstrap approach derived by repeating the analysis 100 times from 300 random points per image from the original data set from the 14 July 2000 CME. The x-axis represents the number of occurrences at a particular value.

Wang-Sheely-Arge

The Wang-Sheely-Arge (WSA) model is then used to calculate the characteristics of the solar wind for input into ENLIL. WSA is an empirical and physics-based model that is used to predict interplanetary magnetic field (IMF) polarity at Earth and the

background solar wind speed which are used to determine the inner boundary conditions for ENLIL. This model uses solar magnetogram measurements as inputs in order to make the calculations. The model then computes the solar wind speed using an empirical relationship that is based upon the divergence of the magnetic field and how close the selected open field lines are to the nearest coronal hole boundary. Additional details on the WSA model can be found in the definitive work from Arge and Pizzo [2000].

ENLIL

ENLIL is named after the Sumerian god "Enlil" whose name literally means "Lord of the Storm" and was considered to be the god of wind or sometimes the god of weather in general. It is used to describe the propagation of the solar wind (to include a CME) outward from the Sun and determine if and when the CME will impact the Earth if one was included. ENLIL approximates the time dependent solution to the MHD equations governing from 21.5 solar radii out to the desired limit. The limit in this work is 1.1 AU while looking at the impact of a CME on Earth. In order to simulate the propagation of a CME, ENLIL will take input parameters that are calculated from the Coned model and the boundary conditions calculated from the WSA model.

The MHD simulation method that is used here involves solving a set of partial differential equations based on the ideal MHD equations. These equations, in metric units are given by

$$\frac{\partial}{\partial t}(\rho) + \nabla \cdot (\rho\, \mathbf{V}) = 0, \tag{9}$$

$$\frac{\partial}{\partial t}(\rho\, \mathbf{V}) + \nabla \cdot (\rho\, \mathbf{VV}) = -\nabla(P) + \nabla \cdot \left(\frac{\mathbf{BB}}{\mu}\right) + \frac{\rho G M_{sun}}{r^2},$$

$$\frac{\partial}{\partial t}(E) + \nabla \cdot (E\mathbf{V}) = -p\nabla \cdot (\mathbf{V}),$$

$$\frac{\partial}{\partial t}(\boldsymbol{B}) = \nabla \times (\mathbf{V} \times \boldsymbol{B}),$$

where **V** is the average flow velocity, ρ is the total mass density, p is the thermal pressure, **B** is the magnetic field, G is the gravitational constant, M_{sun} is the mass of the sun, P is the sum of the thermal pressure and the magnetic ($B^2/2\mu$) pressure, μ is the permeability, $E = p/(\gamma\text{-}1)$ is the thermal energy density, and $\gamma = 5/3$ is given as the ratio of specific heats [Odstrcil and Pizzo, 1999]. Simultaneously, two additional continuity equations must be solved to conserve mass and the magnetic field polarity injected by the CME:

$$\frac{\partial}{\partial t}(\rho_c) + \nabla \cdot (\rho_c \mathbf{V}) = 0, \tag{10}$$

$$\frac{\partial}{\partial t}(\rho_p) + \nabla \cdot (\rho_p \mathbf{V}) = 0,$$

where ρ_c is the density of the injected CME material and ρ_p is the density of the magnetic field polarity [Odstrcil and Pizzo, 1999].

The current version of ENLIL assumes that there is no internal magnetic field structure to the CME while allowing the CME propagation to distort the structure of the IMF. The time-dependent solution to the MHD equations describes the motion of the plasma that makes up the CME as well as what effect the CME will have on the IMF and ambient solar wind.

Previous ENLIL with Coned Model Analyses

The utilization of the Cone model to determine the input parameters of a CME, using the WSA model to determine the boundary conditions for the ambient solar wind structure, and using ENLIL to solve the MHD equations seemed to become the basis for CME modeling after the study of the 12 May 1997 CME by Odstrcil et al. [2005]. This simulation found that it was becoming more feasible to simulate the ambient solar wind parameters and large scale structures in order to estimate the propagation times of CMEs to Earth.

Taktakishvili et al. [2011] used the WSA-ENLIL with Cone model to analyze 36 CMEs which caused large geomagnetic storms, $Kp \geq 8$, using both the analytical Cone model developed by Xie et al. [2004] and the automatic Coned model developed by Pulkkinen et al. [2010] in order to determine the cone parameters for input into ENLIL. The median values of the cone parameters calculated from the Coned model were used as the inputs in the second case. The mean absolute propagation time forecast error for the analytical method was found to be 6.9 hours. The mean absolute propagation time forecast error for the automatic Coned model, method was found to be 11.2 hours. The predicted Kp index in both methods was found to be overestimated. This analysis showed that, while the Coned model was not yet as good as the analytical Cone model, it could be used in order to predict the arrival time and Kp index of CMEs with large geomagnetic storms by a more automated method than before.

Both the analytic Cone model and the Coned model version 1.2 were used later to analyze the propagation of CMEs to Earth and Mars [Falkenberg et al., 2011]. The study found that both the velocity and width of the CME were underestimated by the Coned

model version 1.2 which led to the creation of version 1.3 of the Coned model. Coned model version 1.3 added a modification to the optimization routine to increase the velocity and width estimation to better match the observations and cone parameters predicted by the analytic Cone model.

The Coned model version 1.3 was then compared to the previous version with 15 CMEs [Emmons, 2012]. The Coned model version 1.2 was found to have a mean absolute forecast error of 13.8 hours. The Coned model version 1.3 was found to have a mean absolute forecast error of 9.1 hours. While this did show a great improvement over the previous version, Emmons [2012] found that Coned model tended to push the propagation axis of the CME along the Earth-Sun line. It was suggested that if a weighting factor was introduced for the CME location, that more accurate results might be obtained.

III. Methodology

This chapter discusses the methodology used for the ensemble forecasting of CMEs using WSA-ENLIL with the Coned model. The optimizations performed on the Coned model in order to improve the predictions of the propagation time of CMEs are discussed. Then the core analysis is described as well as the additional analyses completed to analyze the performance of the improved Coned model in relation to previous versions. The changes made to the Coned model are listed as well as the analysis of the model results. The procedure used for determining the actual propagation times is also discussed. Next, the possibility of removing the climatological weighting of the CME cone opening angle is examined. Finally, the effect that a CME can have on the next CME is examined.

Optimizing the Coned Model

Three changes were made in the Coned model in order to improve the predictive ability of the propagation time as well as making the model easier to use. First, the CME threshold value was simplified. Second, additional images were used to determine the initial conditions of the CME. Next, a weighting was added to push the propagation axis of the CME to be more radially outward from the flare location.

The threshold that the Coned model uses to determine what part of the image is the CME and what part is the background is user selectable. Previously, Emmons [2012] varied the value of this input slightly from 56% to 60% of the maximum brightness on the image. Varying this value can improve on the forecasting abilities of the WSA-ENLIL with Coned model; however, each change requires careful observation of the

output after the binary image is generated. Even then, the changes from 56% to 60% are minor and any improvement or worsening of the prediction can only be determined *a posteriori*. For the sake of simplicity, the value of 56% was chosen as this was the most common value used by Emmons [2012] and because the threshold level of 56% was found to be the optimal level for most CMEs by Pulkkinen [2012].

Next, the number of LASCO C3 difference images used was increased. The Coned model allows for any number of images greater than two to be used, however, only three images have been used up to now [Emmons, 2012; Pulkkinen et al., 2010; etc.]. In order to provide additional data to determine the initial conditions of the CMEs, four images were used in all cases except the 28 October 2003 and 3 April 2010 CMEs where four images were not available.

Finally, a weight was added in order to push the propagation axis towards the flare location. Previously, it was discovered that the Coned model pushed the propagation axis of the CME towards the Earth-Sun line [Emmons, 2012]. A weighting was added to the Coned model utilizing the flare location. This pushed the propagation axis closer to radially outward from the flare location without forcing the propagation axis to be exactly radially outward from the flare location.

In order to choose the best weighting for the associated flare location, test runs were done to compare the impact different weights had on the resultant CME arrival time predictions. The weightings of 3×10^{11}, 5×10^{11}, 7.5×10^{11}, 1×10^{12}, 2.5×10^{12}, and 5×10^{12} were chosen since they were near the maximum weighting of 3×10^{11} that was currently used in a run for the climatological value of the opening half angle of the CME.

After all runs were tested and compared, 5×10^{11} was found to be the best weighting factor and that was used in all full ensemble runs.

Core Analysis

For the core analysis, an ensemble forecast was calculated for the CMEs using the WSA-ENLIL version 2.7 with Coned model version 1.4. For each CME, the Coned model used LASCO C3 difference images of the CME eruption to generate a distribution of the initial states of the CME and produced 100 sets of initial conditions. These 100 sets of initial conditions were then used as an input into ENLIL in order to obtain the ensemble forecast distributions.

The two results produced from the ensemble forecast distributions were the propagation time of the CME to the Earth and the maximum Kp index due to the CME impact on the Earth's magnetosphere. The changes in the Coned model were designed to improve upon the propagation time. The Kp index was analyzed and no difference was observed from version 1.3. Therefore, no additional analysis nor tests were performed on the Kp values.

The Coned model version 1.4 was used to produce 100 sets of input parameters for each CME. Each set of these input parameters included the CME velocity, the cone angular width, as well as the latitude and longitude of the propagation axis of the CME cone. The Coned model randomly selected 300 points inside the location of the CME mass in each LASCO C3 difference image and then used these 300 points to calculate the four input parameters. This process was repeated 100 times in order to obtain 100 sets of input parameters. All sets of input parameters were optimized solutions to Equation (11),

$$\min_{\{\theta,\omega,x_0,v\}}\left[\sum_{i}^{N}\sqrt{(\hat{y}_i'-y_i')^2+(\hat{z}_i'-z_i')^2}+\mu|\omega-\omega_0|\right.$$
$$\left.+\beta|\lambda-\lambda_0|+\beta|\phi-\phi_0|\right], \tag{11}$$

where Equation (11) is the updated version of Equation (6) where the new weighting for the flare location is β, λ is the latitude of the calculated propagation axis, λ_0 is the latitude of the flare, ϕ is the longitude of the calculated propagation axis, and ϕ_0 is the longitude of the flare.

These 100 sets of input parameters were then entered into ENLIL in order to calculate the future state of the CMEs at Earth. The other ENLIL parameters were all held constant during the forecasts so that the only variation in predictions were due to variations in the input parameters calculated by the Coned model. Each set of outputs from ENLIL provided a propagation time to Earth and a worst-case maximum Kp index.

The calculated propagation times were compared to the actual propagation times. The propagation times for the original 15 CMEs were taken from Emmons [2012] and compared with the arrival times logged in the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center's (SWPC) historical weekly reports (http://www.swpc.noaa.gov/ftpmenu/warehouse.html) and with the ACE data from NASA's OMNIWeb database (http://ftpbrowser.gsfc.nasa.gov/ace_merge.html) where the arrival times were determined by a sharp increase in the magnetic field magnitude, solar wind speed, and solar wind particle density in the solar wind measurements. For the final six CME measurements, the NOAA/SWPC historical

weekly reports were used and checked against the ACE data to determine the impact times.

The associated solar flare locations were taken from the NOAA/SWPC historical weekly reports and were used to approximate the locations of the CME eruptions. The actual measured values for the propagation time, maximum Kp indices, and locations of the solar flares are displayed in Table 3 for the original 15 CMEs done in the work by Emmons [2012]. The actual measured values for the propagation time, maximum Kp indices, and locations of the solar flares are displayed in Table 4 for the additional 6 CMEs done to check the validity of the new Coned model version 1.4.

Table 3: The start date and times, actual propagation times as measured by ACE, maximum Kp indices as measured for the 15 original CMEs analyzed with CME event number labeled for reference.

Event Number	CME Start Date (YYYMMDD)	CME Start Time (UT)	Propagation Time to ACE (hours)	Maximum Kp	Associated Solar Flare Location
1	19990503	06:06	56.83	3	N15E32
2	20000404	16:32	47.50	9	N16W66
3	20000714	10:54	27.33	9	N22W07
4	20010329	10:26	37.83	9	N20W09
5	20010410	05:30	33.83	8	S23W09
6	20010924	10:30	33.50	7	S16E23
7	20011009	11:30	52.75	6	S28E08
8	20011104	16:35	32.67	9	N06W18
9	20011117	05:30	60.00	4	S13E42
10	20031028	11:30	18.33	9	S16E08
11	20031029	20:54	19.83	9	S15W02
12	20040720	13:31	44.33	7	N10E35
13	20041106	02:06	39.67	9	N07E00
14	20041203	00:26	54.33	4	N09E03
15	20100403	10:34	45.25	8	S25E00

The ensembles were run on a dual core 2.93 GHz Intel machine which required approximately 3 days to complete each full 100-run ensemble. Since each run can be done independently, this can be designed to be calculated in parallel in order to vastly reduce the computational time necessary to run the ensemble.

Table 4: The start date and times, actual propagation times as measured by ACE, maximum Kp indices as measured for the 6 additional CMEs analyzed with CME event number labeled for reference.

Event Number	CME Start Date (YYYMMDD)	CME Start Time (UT)	Propagation Time to ACE (hours)	Maximum Kp	Associated Solar Flare Location
16	19980502	14:06	36.38	9	S15W15
17	20000809	16:30	53.25	8	N14W66
18	20011019	16:50	47.40	8	N15W29
19	20011122	23:30	30.15	8	S15W34
20	20031118	08:50	46.83	9	N00E18
21	20061213	02:54	35.03	8	S06W24

Model Input

The first step in producing an ensemble forecast using WSA-ENLIL with Coned model is to run the Coned model for a particular event. The Coned model requires a series of LASCO C3 images of the CME eruption in order to calculate the ensemble of input parameters. These images were found at the Community Coordinated Modeling Center's (CCMC) iNtegrated Space Weather Analysis System (iSWSA) located at http://iswa.gsfc.nasa.gov/IswaSystemWebApp/. This analysis used four images for each CME except for CMEs number 10 and 15 (the 2003-10-28 and 2010-04-03 CMEs respectively) since four good images of the CME eruption could not be found. For the two CMEs where four images were unavailable, the same three images used by Emmons

[2012] were used. The Coned model also contains a threshold level for filtering the images to determine the location of the CME mass by analyzing the brightness of each pixel in the provided images. The brightest pixels in the image correspond to the location of the CME plasma. This threshold level is the percentage of the normalized intensity used to select the CME mass from the LASCO C3 images provided. The threshold level ranges from zero to one with zero selecting everything in the image and one selecting nothing in the image. The time stamps of the LASCO images along with the associated solar flare location are used as input to the Coned model along with the filtering threshold level. The time stamps used can be seen in Table 5 while the associated solar flare locations used are given in Table 3 and Table 4 above.

Table 5: The list of the time stamps of the LASCO C3 images used as inputs to the Coned model by event number and CME start date.

Event Number	CME Start Date (YYYMMDD)	LASCO C3 Image Time Stamps ('YYYMMDDHHMMSS')			
1	19990503	'19990503074200'	'19990503081800'	'19990503084200'	'19990503091800'
2	20000404	'20000404164300'	'20000404171800'	'20000404174200'	'20000404181800'
3	20000714	'20000714111800'	'20000714114200'	'20000714121800'	'20000714124700'
4	20010329	'20010329114200'	'20010329121800'	'20010329124200'	'20010329134200'
5	20010410	'20010410061800'	'20010410064200'	'20010410074200'	'20010410081800'
6	20010924	'20010924111800'	'20010924114200'	'20010924121800'	'20010924124200'
7	20011009	'20011009121800'	'20011009124200'	'20011009134200'	'20011009141800'
8	20011104	'20011104170000'	'20011104173000'	'20011104180200'	'20011104185400'
9	20011117	'20011117074200'	'20011117084200'	'20011117094200'	'20011117102300'
10	20031028	'20031028114200'	'20031028121800'	'20031028124200'	
11	20031029	'20031029214200'	'20031029221800'	'20031029231800'	'20031029234200'
12	20040720	'20040720151800'	'20040720154200'	'20040720161800'	'20040720164200'
13	20041106	'20041106021800'	'20041106024200'	'20041106041800'	'20041106051800'
14	20041203	'20041203014200'	'20041203021800'	'20041203024200'	'20041203031800'
15	20100403	'20100403114200'	'20100403121800'	'20100403134200'	
16	19980502	'19980502154200'	'19980502164200'	'19980502174600'	'19980502184500'
17	20000809	'20000809181800'	'20000809184200'	'20000809194200'	'20000809201800'
18	20011019	'20011019181800'	'20011019191100'	'20011019194200'	'20011019201800'
19	20011122	'20011123014200'	'20011123021800'	'20011123024200'	'20011123031800'
20	20031118	'20031118104200'	'20031118111800'	'20031118114200'	'20031118121800'
21	20061213	'20061213031800'	'20061213034200'	'20061213041800'	'20061213044200'

After the Coned model run is completed, 100 sets of input parameters are created and put into a separate control file for each set. The Coned model requires about 5 minutes to complete on a desktop computer using an AMD Athlon 7750 dual core processor at 2.70 GHz with 4 GB of RAM. These control files can then be used as inputs into ENLIL.

In order to run WSA-ENLIL, WSA must first be run for the appropriate Carrington rotation date and the solar wind and IMF solution are used as the inner boundary conditions for ENLIL. The input parameters for ENLIL were all held constant except for the Coned model outputs of the CME velocity, angular width, and axis of propagation (Table 6). Magnetogram measurements were available from multiple source locations but in order to match up with the work of Emmons [2012], the magnetograms measured by the Kitt Peak National Observatory was used for all CMEs. The low resolution (160x30x90) option for the ENLIL computational grid was used for all CMEs due to the large computation required for the high resolution runs as well as to match up with the previous work of Emmons [2012].

Table 6: A list of the input parameters for the WSA-ENLIL with Coned model along with their default values.

Input Parameter	Value
Magnetogram Source	NSO-Kitt Peak
Number of Cone Clouds	1
Outer Radial Boundary	1.1 AU
Fast Stream Solar Wind Density	200 cm^{-3}
Fast Stream Solar Wind Temperature	0.8×10^6 K
Fast Stream Solar Wind Speed	625 km/s
Fast Stream Radial Magnetic Field	300 nT
Minimum Solar Wind Speed	225 km/s
Magnetic Field Scaling Factor	2.5 (for NSO-Kitt Peak)
Fraction of Alpha Particles to Protons	0.03
Cloud Start Date	Variable
Cloud Start Time	Variable
Latitude of Cloud Center	Variable
Longitude of Cloud Center	Variable
Radius of Cloud	Variable
Cloud Velocity	Variable
Density Enhancement Factor	4
Temperature Enhancement Factor	1
Elongation Factor	1
Shape of Cloud	Spherical
Resolution	160x30x90

Analysis of Model Output

The output from WSA-ENLIL with Coned model was analyzed to determine the propagation time to the Earth as well as the maximum Kp index. The arrival time of the CME at Earth was selected to be the time given from the NOAA/SWPC historical weekly reports which was confirmed by finding the time of the sharp increase in the magnetic field magnitude, solar wind speed, and solar wind particle density in the solar wind measurements from ACE data. This propagation time was then compared to the propagation time calculated from the outputs from WSA-ENLIL with Coned model.

The actual maximum Kp indices were taken from NASA's OMNIWeb database. These values were then compared to the maximum possible Kp index values calculated from the outputs from WSA-ENLIL with Coned model assuming a due south IMF. Assuming a due south IMF gives the maximum possible Kp index but will be an overestimation in the case that the IMF is not due south. This assumption was found to overestimate the Kp index in general [Emmons, 2012].

In order to analyze the ensemble distributions, various statistical calculations were performed on the propagation times, maximum Kp indices, and the input parameters. These calculations included the average, standard deviation, median, median absolute deviation, and range as well as determining the minimum and maximum values. The forecast error was also calculated for the propagation time and the Kp by comparing the average and median values of the ensemble forecast distributions to the actual values. In addition, the mean absolute error was calculated for the propagation time and maximum Kp.

These statistics were then compared with the results from the Coned model version 1.3 [Emmons, 2012]. The mean absolute difference as well as the percentage of improvement in the mean absolute error was calculated. Additionally, the improvement was determined in the number of CME predictions that fell within the range, one median absolute deviation, and one, two, and three standard deviations.

Removal of Climatological Weighting

During this analysis, it was noticed that with the addition of the CME eruption location, there was more actual data available to the Coned model to calculate the input

parameters. The climatological weighting of the CME cone opening angle which was added in order to stabilize the solution in some cases [Pulkkinen et al., 2010] might not be needed anymore. To test the effects of the removal of this climatological weight, the original 15 CMEs were recalculated using the single-shot method and compared to the results of those CMEs with the weighting included.

Multiple CMEs in WSA-ENLIL with Coned Model

Finally, the effect of multiple CMEs occurring within a short period of time was analyzed. CMEs slow down from their impact with the background solar wind [Gopalswamy et al., 2000]. When multiple CMEs happen near each other, an earlier CME can "clear out" a path for a later CME [Skoug et al., 2004]. Therefore, a set of multiple CMEs was run together in order to examine the effects that they would have on each other and see if this could improve upon the results.

IV. Results

This chapter begins with the results from the Coned model version 1.4 with weighting for the flare location and four images for the original 15 CMEs compared to the original Coned model version 1.3 results. The 6 additional CMEs are then presented in order to verify the validity of the improvements to the Coned model. Next, additional options are analyzed to determine if the old climatological weighting of the opening angle of the cone can be safely removed. Finally, multiple CME Coned model inputs were run through ENLIL together in order to see if this can improve cases where multiple CMEs happen over a short period of time.

Coned model version 1.4

Input Parameters

The input parameters calculated by the Coned model are the cone opening half angle, the velocity of the cone front, and the latitude and longitude of the propagation axis of the CME cone. The distribution of the initial states for the first 15 CMEs, as calculated by the Coned model version 1.4, is displayed in Table 7 and Table 8. The full set of values of the ensemble input parameters calculated from the Coned model and the filtered LASCO C3 difference images used are left out for brevity but are available upon request.

The added weight for the CME location into the Coned model version 1.4 successfully pushed the propagation axis of the CME cone towards the flare location for these 15 CMEs. For the latitudes, the original Coned model version 1.3 averaged 3.80° away from the Earth-Sun line. This is small compared to the average of 16.07° away

from the solar equator for the latitude of the associated flare location. The Coned model version 1.4 increased the average distance away from the Sun-Earth line by 24% to 4.71°. The original Coned model version 1.3 only had a single occurrence where the propagation latitude varied more than 10° from the Earth-Sun line. This was the 10 April 2001 CME. After the improvements were made, there were three cases where the calculated latitude was greater than 10° from the Earth-Sun line (3 May 1999, 10 April 2001, and 9 October 2001).

Table 7: Statistics for the input latitude distribution of the initial 15 CMEs derived from the Coned model version 1.4. A negative angle represents a southward direction while a positive angle represents a northward direction.

CME date (YYYYMMDD)	average (deg)	standard deviation (deg)	median (deg)	median absolute deviation (deg)	range (deg)	minimum (deg)	maximum (deg)
19990503	11.19	3.31	12.00	2.00	11.00	4.00	15.00
20000404	0.73	0.60	1.00	0.00	2.00	0.00	2.00
20000714	3.37	1.02	3.00	1.00	5.00	2.00	7.00
20010329	-0.04	0.20	0.00	0.00	1.00	-1.00	0.00
20010410	-12.67	3.27	-12.00	2.00	13.00	-20.00	-7.00
20010924	-7.36	1.14	-7.00	1.00	4.00	-9.00	-5.00
20011009	-10.53	2.77	-10.00	2.00	15.00	-20.00	-5.00
20011104	-0.15	0.36	0.00	0.00	1.00	-1.00	0.00
20011117	3.59	1.14	3.00	1.00	4.00	2.00	6.00
20031028	0.09	0.29	0.00	0.00	1.00	0.00	1.00
20031029	-3.33	0.87	-3.00	1.00	4.00	-6.00	-2.00
20040720	0.15	0.36	0.00	0.00	1.00	0.00	1.00
20041106	6.18	0.86	6.00	1.00	4.00	4.00	8.00
20041203	8.54	0.58	9.00	0.00	3.00	7.00	10.00
20100403	-2.75	1.01	-3.00	1.00	4.00	-5.00	-1.00

The longitudes were changed even more than the latitudes were by the improvements made to the Coned model. For these 15 CMEs, the original Coned model

version 1.3 averaged 4.97° longitude away from the Earth-Sun. Again, this is much smaller than the average of 18.13° away from the solar equator for the longitude of the associated flare location. The Coned model version 1.4 increased the average distance away from the Sun-Earth line by 44% to 7.13° longitude. The original Coned model version 1.3 only had three instances where the propagation longitude varied more than 10° from the Earth-Sun line (3 May 1999, 24 September 2001, and 17 November 2001). After the improvements were made, there were four cases where the calculated longitude was greater than 10° from the Earth-Sun line (3 May 1999, 4 April 2000, 24 September 2001, and 17 November 2001).

Table 8: Statistics for the input longitude distribution of the initial 15 CMEs derived from the Coned model version 1.4. A negative longitude represents an eastward direction while a positive longitude represents a westward direction.

CME date (YYYYMMDD)	average (deg)	standard deviation (deg)	median (deg)	median absolute deviation (deg)	range (deg)	minimum (deg)	maximum (deg)
19990503	-24.05	7.23	-26.00	5.00	26.00	-33.00	-7.00
20000404	18.02	6.43	16.00	4.00	23.00	8.00	31.00
20000714	5.41	1.07	5.00	1.00	4.00	3.00	7.00
20010329	0.03	0.17	0.00	0.00	1.00	0.00	1.00
20010410	2.01	0.77	2.00	0.00	3.00	1.00	4.00
20010924	-20.73	2.80	-22.00	1.00	11.00	-24.00	-13.00
20011009	2.78	0.73	3.00	1.00	4.00	1.00	5.00
20011104	7.65	1.89	7.00	1.00	8.00	4.00	12.00
20011117	-16.56	5.19	-16.00	4.00	19.00	-26.00	-7.00
20031028	0.03	0.17	0.00	0.00	1.00	0.00	1.00
20031029	1.96	0.53	2.00	0.00	3.00	1.00	4.00
20040720	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20041106	-2.39	0.53	-2.00	0.00	2.00	-3.00	-1.00
20041203	-3.08	0.27	-3.00	0.00	1.00	-4.00	-3.00
20100403	2.32	0.80	2.00	1.00	3.00	1.00	4.00

Overall, the additional weighting for the location of the flare did what it was expected to do. Push the propagation axis towards the flare location while not forcing it to match exactly. A comparison between the calculated latitudes for the original Coned model version 1.3 and the updated version with the latitude of the flare location is given in Figure 9. Additionally, a comparison between the calculated longitudes for the Coned model version 1.3 and Coned model version 1.4 is given in Figure 10.

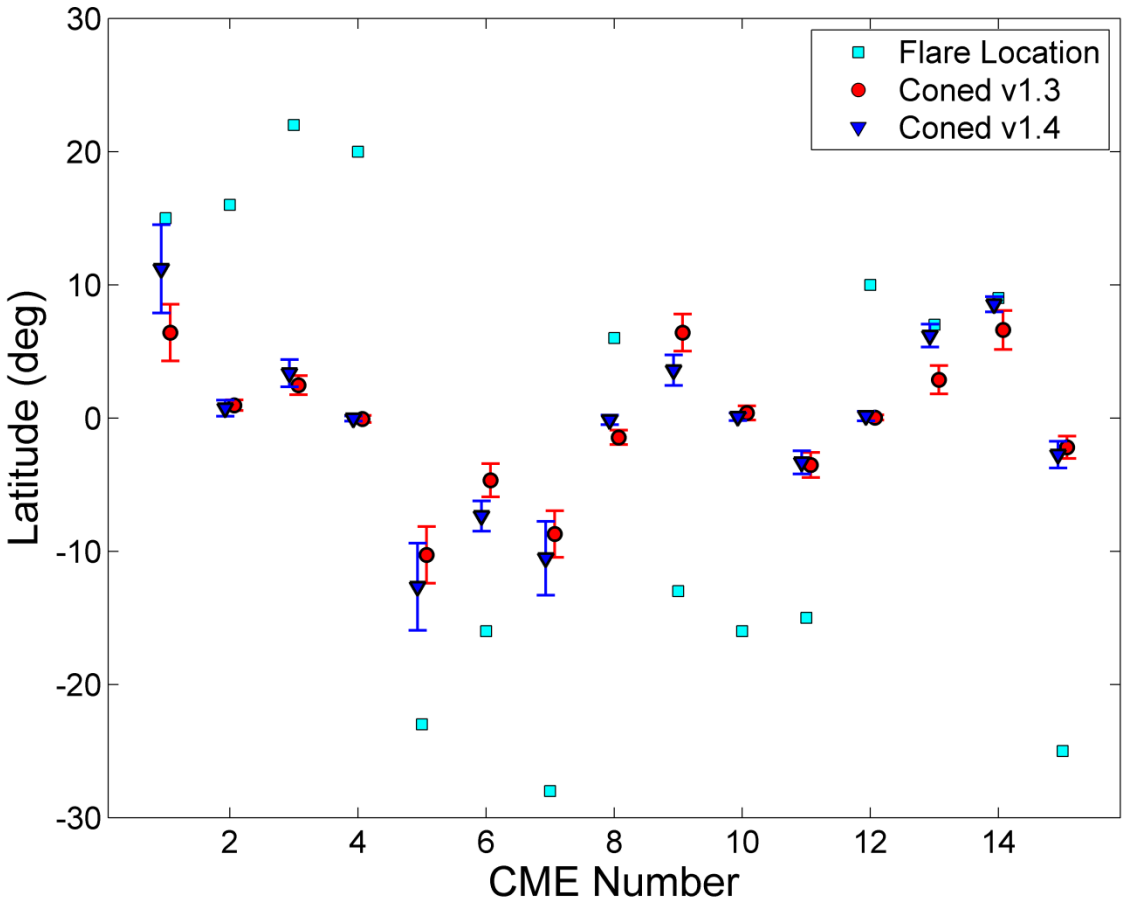

Figure 9: Comparison between the calculated cone latitude for the Coned model version 1.4 and the original Coned model version 1.3 with the flare location noted for reference. The symbols are offset slightly to allow differentiation between values for the same CME.

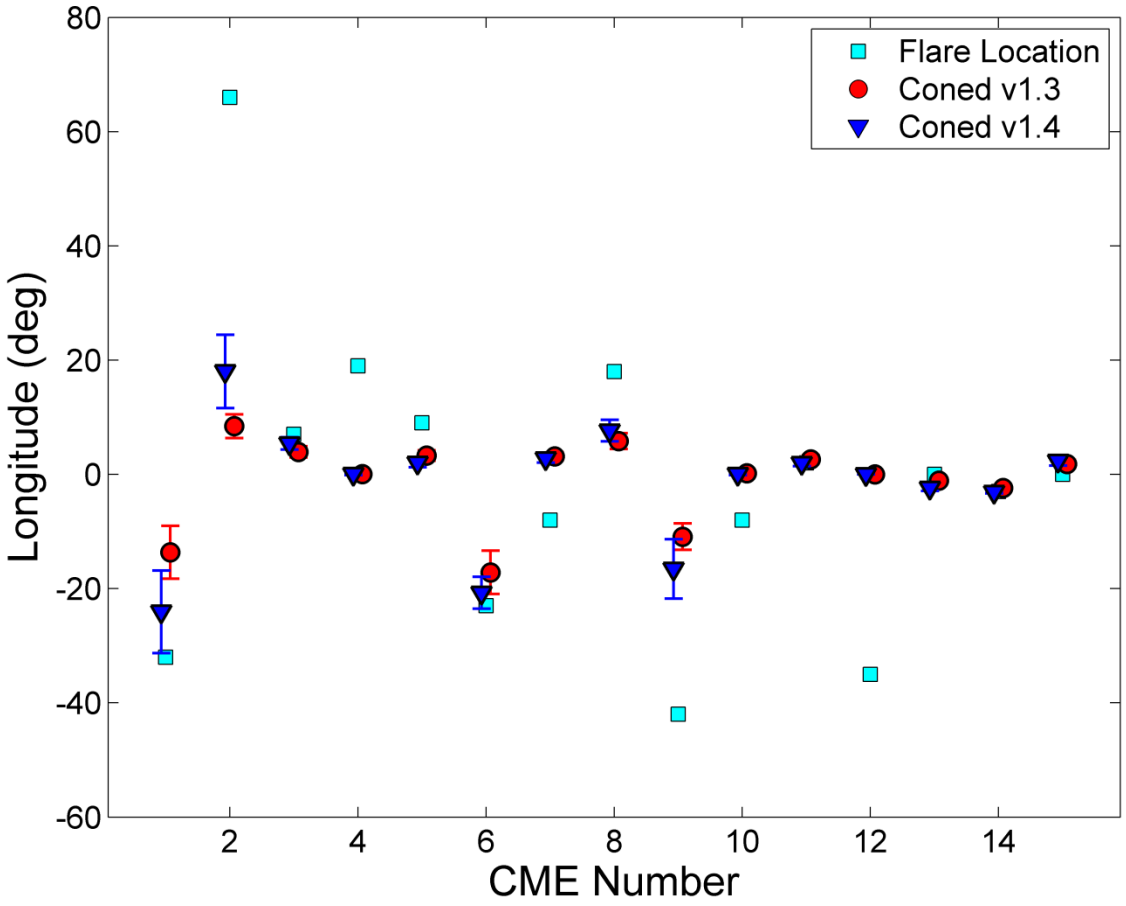

Figure 10: Comparison between the calculated cone longitude for the Coned model version 1.4 and the original Coned model version 1.3 with the flare location noted for reference. The symbols are offset slightly to allow differentiation between values for the same CME.

Despite the original Coned model version 1.3 clustering the propagation axis along the Earth-Sun line, there was a positive correlation between the latitude of the solar flare and the latitude of the calculated CME cone propagation. A common method of determining this correlation is the correlation coefficient (Pearson's) which describes the degree of linear dependence between two data sets. A correlation coefficient greater than 0.5 is commonly interpreted as a strong correlation. A p-value is used to describe the probability that the correlation occurred by chance and that randomly selected points could have the same relationship. A p-value less than 0.05, i.e. a 5% probability the

correlation occurred by chance, is commonly accepted as the criterion for a statistically significant correlation. The original Coned model version 1.3 latitudes had a correlation coefficient of 0.63 and a p-value of 0.01 with the latitude of the flare location while the longitudes had a correlation coefficient of 0.73 and a p-value of 0.00. The Coned model version 1.4 increased this correlation coefficient to 0.70 with a p-value of 0.00 for the latitude and increased the correlation coefficient to 0.80 with a p-value of 0.00 for the longitude. This positive correlation between the solar flare location and the calculated latitude of the CME cone propagation is shown in Figure 11 for the Coned model v1.4.

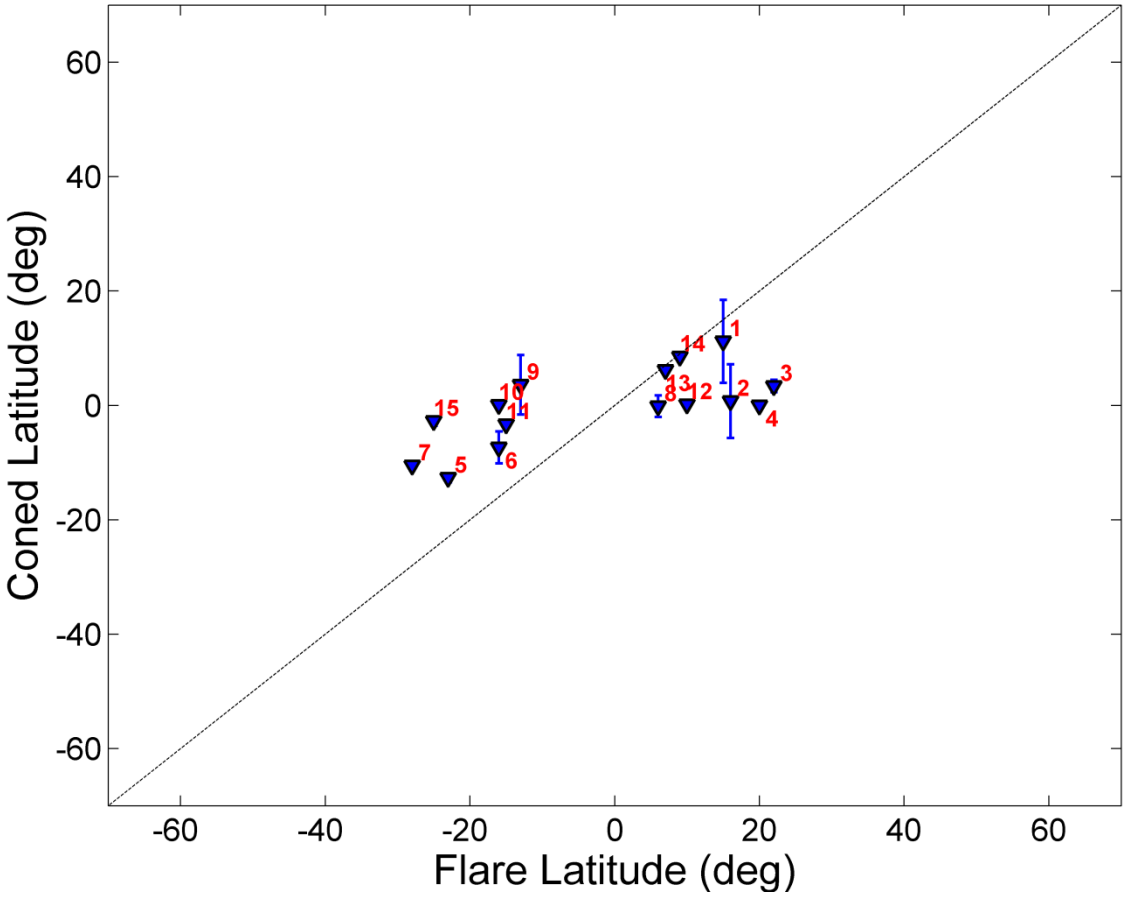

Figure 11: Comparison between the calculated cone latitude for the Coned model version 1.4 and the flare location to show correlation.

The correlation between the solar flare location and the calculated longitude of the CME cone propagation is shown in Figure 12. The CME run number is labeled for each run in both figures. The outlier of CME 2 in Figure 12 is because the CME occurred at the edge of the solar disk but was directed towards the Earth.

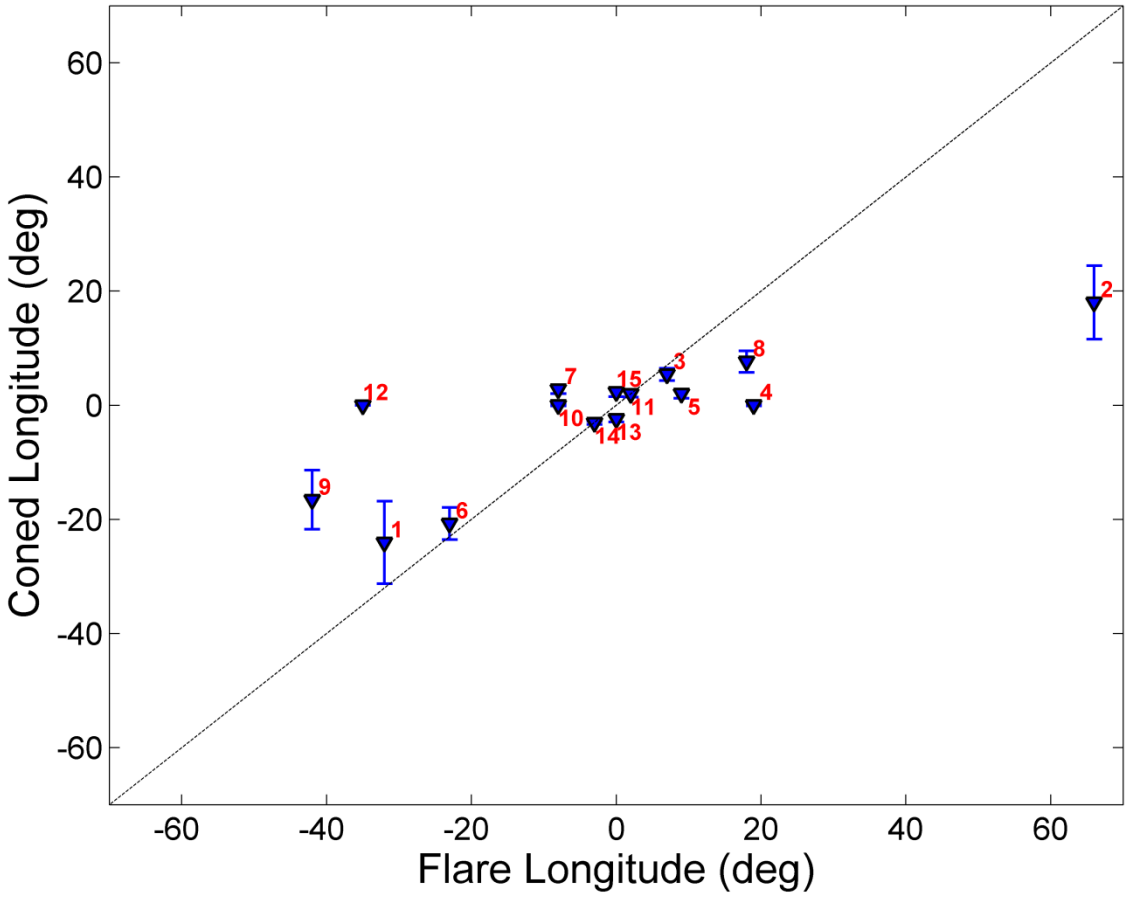

Figure 12: Comparison between the calculated cone longitude for the Coned model version 1.4 and the flare location to show correlation.

Propagation Time

The improvements made to the Coned model version 1.3 greatly increased the accuracy of the forecasts for these 15 CMEs. Emmons [2012] noted was that the average of the ensemble averages of the propagation times for the 15 CMEs was 36.7 hours with a standard deviation of 7.1 hours for the Coned model version 1.3. This showed a much smaller spread when compared to the actual average of the 15 CMEs of 40.3 hours with a standard deviation of 12.9 hours. The average of the ensemble averages for the improved version was calculated to be 40.4 hours with a standard deviation of 8.5 hours (Table 9). This shows that the Coned model version 1.4 now has the average propagation time centered much closer to the correct time while having a more accurate distribution of results that is not as centered upon the average propagation time.

Table 9: The statistics for the propagation time for the original 15 CMEs using the Coned model version 1.4.

CME date (YYYYMMDD)	actual (hours)	average (hours)	standard deviation (hours)	median (hours)	median absolute deviation (hours)	range (hours)	minimum (hours)	maximum (hours)
19990503	56.83	54.49	8.52	57.85	4.44	35.05	31.10	66.15
20000404	47.50	46.11	8.38	44.32	6.65	31.55	31.77	63.32
20000714	27.33	33.33	4.31	32.55	2.40	21.92	25.67	47.58
20010329	37.83	37.72	5.21	37.19	3.73	28.45	29.67	58.12
20010410	33.83	41.61	7.05	39.74	4.80	28.23	29.68	57.92
20010924	33.50	34.11	3.21	35.05	1.95	14.68	25.28	39.97
20011009	52.75	46.59	6.81	46.00	4.40	29.38	34.72	64.10
20011104	32.67	30.60	4.60	30.28	3.33	19.80	22.05	41.85
20011117	60.00	43.68	8.79	42.76	6.58	33.18	28.38	61.57
20031028	18.33	26.00	3.30	25.18	2.41	13.40	20.62	34.02
20031029	19.83	27.92	3.66	27.59	2.64	14.80	21.95	36.75
20040720	44.33	50.86	4.80	50.15	2.93	23.95	42.50	66.45
20041106	39.67	40.85	3.42	40.28	2.42	17.10	33.95	51.05
20041203	54.33	44.98	3.96	45.08	2.92	16.18	38.22	54.40
20100403	45.25	47.02	6.33	46.04	4.28	29.43	64.08	34.65

The improvements extended beyond the overall averages and standard deviations, though. The absolute average error was reduced by over 43% from 9.06 hours in the original version to 5.16 hours in the improved version (Table 10). Additionally, the absolute average error of the median forecast for each ensemble result showed an even larger improvement of over 45% from 9.17 hours [Emmons, 2012] to 5.00 hours. The number of CMEs falling within one standard deviation increased from 5 to 8 (Figure 13) and the number of CMEs falling within three standard deviations increased from 11 to 15. Finally, the number of CMEs falling within the range of ensemble results also improved from 8 to 13 (Figure 14).

Table 10: The forecast errors and performance metrics for the propagation time of the original 15 CMEs using the Coned model version 1.4. In this table, *avg* stands for average, *med* stands for median, *std* stands for standard deviation, and *mad* stands for median absolute deviation. A negative value represents the predicted arrival time was earlier than the actual time. Improvements over the Coned model version 1.3 are shown in green while red represents a worse result.

CME date (YYYYMMDD)	avg -actual (hours)	actual inside avg ± 1 std?	med-actual (hours)	actual inside med ± 1 mad?	actual inside range?
19990503	-2.34	yes	1.02	yes	yes
20000404	-1.39	yes	-3.18	yes	yes
20000714	6.00	no	5.22	no	yes
20010329	-0.11	yes	-0.64	yes	yes
20010410	7.78	no	5.91	no	yes
20010924	0.61	yes	1.55	yes	yes
20011009	-6.16	yes	-6.75	no	yes
20011104	-2.07	yes	-2.40	yes	yes
20011117	-16.32	no	-17.24	no	yes
20031028	7.67	no	6.85	no	no
20031029	8.09	no	7.76	no	no
20040720	6.53	no	5.82	no	yes
20041106	1.18	yes	0.61	yes	yes
20041203	-9.35	no	-9.26	no	yes
20100403	1.77	yes	0.79	yes	yes
absolute mean	5.16		5.00		

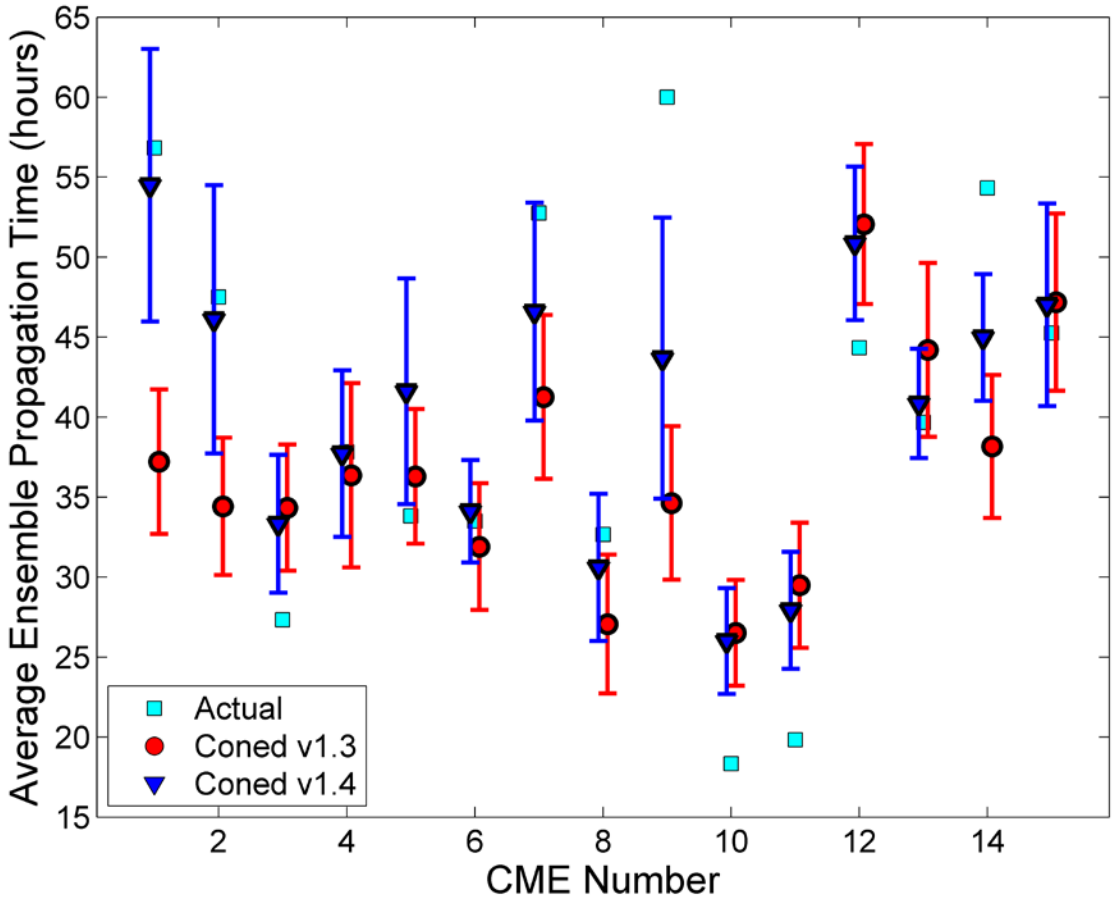

Figure 13: Comparison of the average predicted propagation time with the Coned model version 1.3 (red) and the Coned model version 1.4 (blue). The actual propagation times are given for reference (cyan) and the error bars represent one standard deviation. The symbols are offset slightly to allow differentiation between values for the same CME.

With the improvements, all but two CMEs now fall within the range (28 October 2003 and 29 October 2003). These two CMEs occurred during a particularly active time where several other CMEs were occurring shortly before and after these CMEs. Since these two cases showed two of the three largest positive errors, the simulation is predicting a larger deceleration than was observed. This error could be caused by the fact that only a single CME is analyzed at a time independent of the effects of other CMEs. The background interstellar medium acts to slow down the propagation time and earlier

CMEs will clear out the material in the way as they propagate [Skoug et al., 2004]. It is likely that previous CMEs clearing out this material in the way will cause the CME in question to experience less deceleration and decrease the propagation time. This hypothesis is examined later in this section.

Even including these two cases, out of the 15 CMEs tested, 14 showed improvements while only a single CME was worse. The CME from 10 April 2001 went from an error of 2.46 hours to 7.78 hours. A CME occurred 12 hours prior to this CME and this could have caused the CME to travel faster than ENLIL calculated by looking at the CME independently. The other 14 runs all showed improvements of up to 17 hours. Originally, there were 5 CMEs with an error of greater than 10 hours (Figure 13 and Figure 14), but the Coned model version 1.4 has now reduced that to only a single run worse than 10 hours (17 November 2001). This single run is the only outlier that has more than twice the absolute average error and is 7 hours worse than the next worst run. This is likely due to the irregular shape of this particular CME. The Cone model assumes that the CME can be approximated as a cone. In Figure 15 (a) and (b), two LASCO C3 difference images are shown representative of most of the other CMEs studied here as well as the smooth edge that the Coned model looks for in order to calculate out the parameters for the CME front. The CME for 17 November 2001 is shown in Figure 15 (c). Since the Coned model relies on the assumption that the CME has the shape of a cone, the model was unable to accurately calculate the cone parameters for this case.

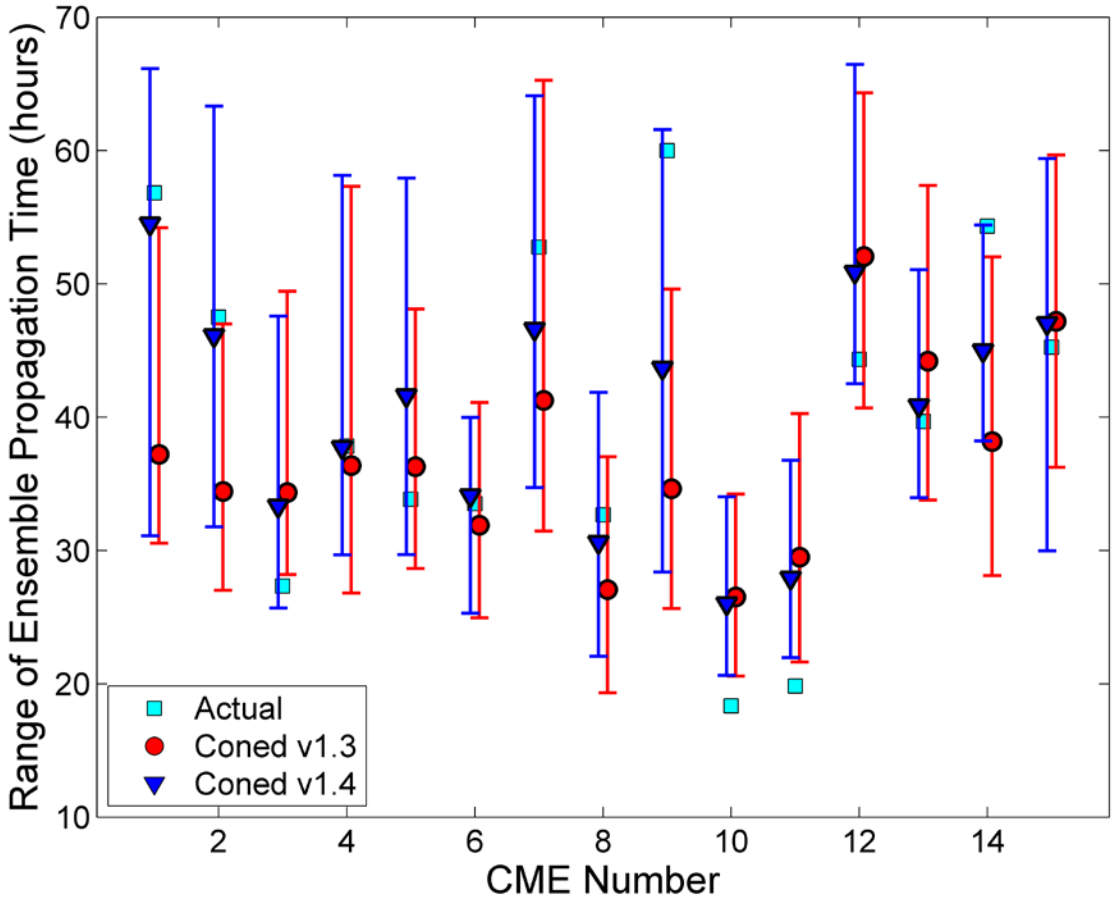

Figure 14: Comparison of the average predicted propagation time of the Coned model version 1.3 (red) and the Coned model version 1.4 (blue). The actual propagation times are given for reference (cyan) and the error bars represent the full range of values calculated in the 100 runs. The symbols are offset slightly to allow differentiation between values for the same CME.

The original Coned model version 1.3 not only performed worse on the forecasts of CME arrival times, but the correlation between the actual propagation time and its prediction was poor. The correlation coefficient for the Coned model version 1.3 predicted propagation time with the actual propagation time was only 0.50 with a p-value of 0.06. This means that the correlation would not be considered a strong correlation, though it is a positive one. The correlation is shown in Figure 16 for the propagation times with the improved Coned model version 1.3. The correlation coefficient was

improved to 0.87 with a p-value of 0.00 representing a very strong and positive correlation between the actual and predicted propagation times.

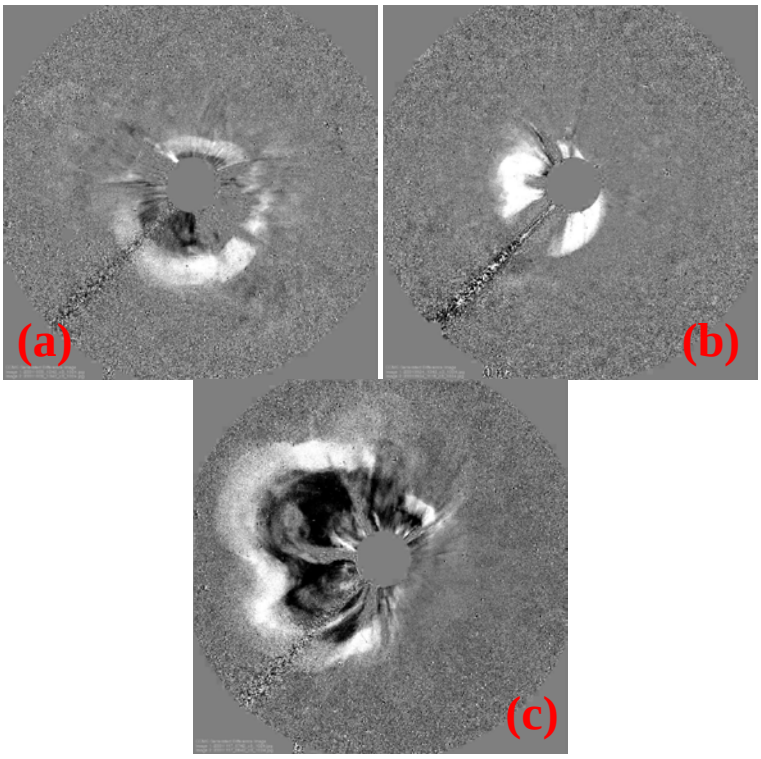

Figure 15: LASCO C3 images showing the difference between the regularity of different CMEs. The CMEs are from (a) 9 October 2001, (b) 24 September 2001, and (c) 17 November 2001.

The correlation coefficient was so poor in the Coned model version 1.3 due, in large part, to the five slowest CMEs. A comparison between the original Coned model version 1.3 and the improved version is given in Figure 17. Originally, all five CMEs that took longer than 46 hours to reach the Earth had their forecasts off by more than 10 hours. These five CMEs had an average absolute error of 17.15 hours. The Coned model version 1.4 improved four out of five of those CMEs to within the 10 hour range. Additionally, the average absolute error of those CMEs was reduced by 59% to 7.11

hours with two out of the five improving to within 2.5 hours of the actual time. The worst runs are still concentrated on the extremes, however, with the two worst results coming from the first and third slowest CMEs and the two fastest CMEs making up the third and fifth worst results. This indicates that there are additional factors causing these extremes to have poor predictions. One possibility is other CMEs happening soon before

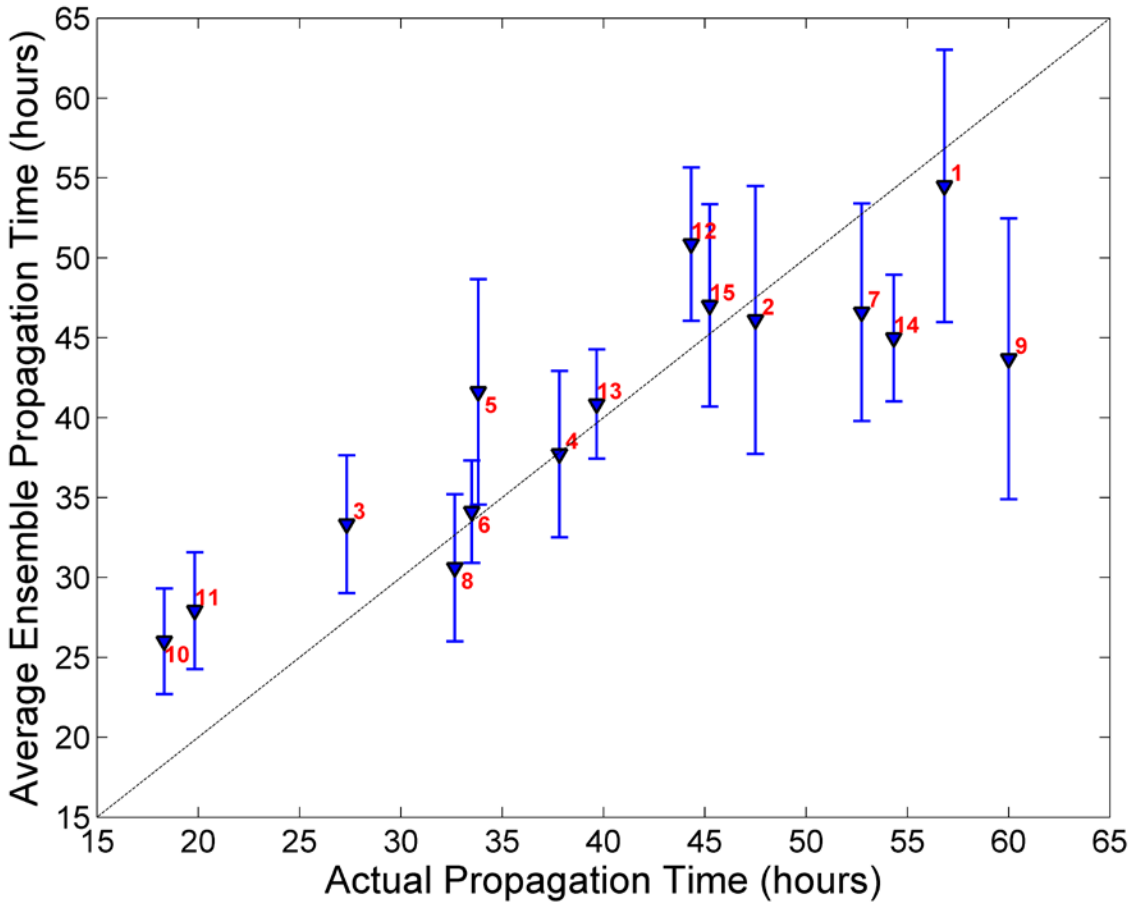

Figure 16: The averages and standard deviations of the ensemble propagation times versus the actual propagation times.

or after the CME changing the makeup of the interstellar medium and thus altering the propagation of the CME. Since WSA-ENLIL with Coned model was run with a single CME at a time, these additional effects were not taken into account for these calculations.

Additionally, CMEs that don't match the expected cone shape are difficult for the Coned model to estimate and this introduces errors in these cases.

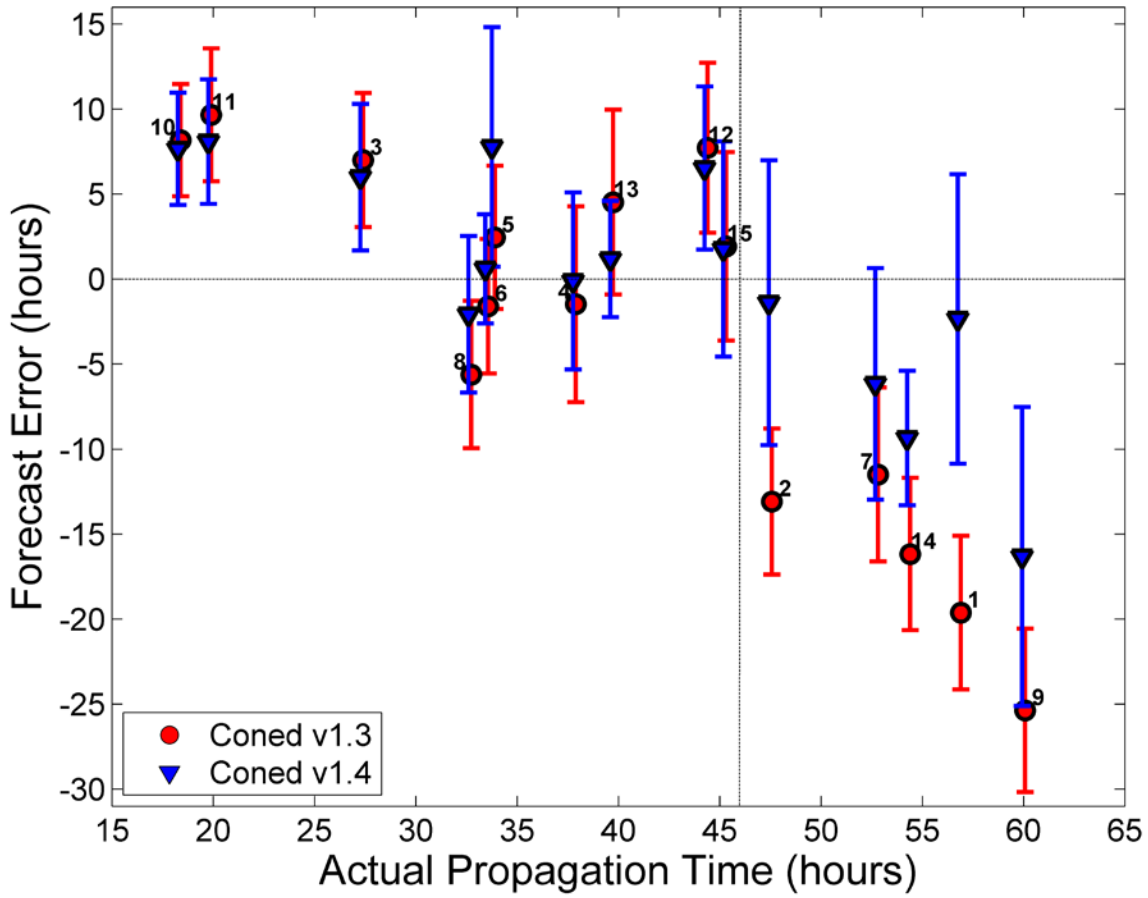

Figure 17: The forecast error for the propagation time versus the actual propagation time. The error bars represent one standard deviation and the vertical line represents the 46 hour point where all CMEs were forecast with an absolute error of more than 10 hours.

Analysis of Additional CMEs

As an additional measure, six extra CMEs were chosen in order to confirm that the improvement in the predictions were due to the enhancements in the methodology and not just specific to those 15 CMEs. The CMEs were chosen randomly from the list of 36 CMEs that caused large geomagnetic storms previously analyzed by Taktakishvili

et al. [2011]. The only requirement on these CMEs was that four images were available. These CMEs were used despite the quality of their images and the occurrence of additional CMEs within a few hours of the main CME.

These 6 CMEs were then run through a single-shot run as well. This was first done with the Coned model version 1.3 and then completed with the updated Coned model version 1.4. The improvements shown in these new CMEs mirrored the improvement of the original 15 CMEs and so these values and weightings were used in order to complete full ensemble runs for the core analysis so that the full statistical information could be determined. These CMEs were listed in Table 4.

Input Parameters

The input parameters turned out much the same way for the extra runs as they did for the original 15 CMEs. In these cases, all six CMEs had the latitude of the associated flare within 15° of the equator so the additional weighting did not change much (Table 11). The correlation is not as useful here since there are too few data points to accurately determine any correlation.

Table 11: Statistics for the input latitude distribution of the six extra CMEs derived from the Coned model version 1.3 (using single-shot runs) and the Coned model version 1.4. A negative latitude represents an southward direction. Here, *std* stands for standard deviation, and *avg* stands for average.

CME date (YYYYMMDD)	actual (deg)	Coned v1.3 avg (deg)	Coned v1.4 avg (deg)	Coned v1.4 std (deg)	Coned v1.4 range (deg)
19980502	-15.00	6.00	6.28	1.50	5.00
20000809	14.00	8.00	9.14	2.13	9.00
20011019	15.00	2.00	2.61	0.96	5.00
20011122	-15.00	0.00	2.16	1.37	8.00
20031118	0.00	-4.00	-5.40	1.28	6.00
20061213	-6.00	-14.00	-10.40	1.85	11.00

While the latitudes of these flares were all close to the solar equator, the longitude of the flare location varied much more. The latitude of these six CMEs were all 15° or greater away from the Sun-Earth line (Table 12). With the larger flare coordinates, the weighting was able to affect the propagation axis more than it did the latitude but the largest change was still under 3° so the changes were kept small in these cases.

Table 12: Statistics for the input longitude distribution of the six extra CMEs derived from the Coned model version 1.3 (using single-shot runs) and the Coned model version 1.4. A negative longitude represents an eastward direction. Here, *std* stands for standard deviation, and *avg* stands for average.

CME date (YYYYMMDD)	actual (deg)	Coned v1.3 avg (deg)	Coned v1.4 avg (deg)	Coned v1.4 std (deg)	Coned v1.4 range (deg)
19980502	15.00	8.00	8.90	2.20	9.00
20000809	66.00	0.00	-1.21	0.41	1.00
20011019	29.00	7.00	9.95	3.51	22.00
20011122	34.00	8.00	10.75	2.16	10.00
20031118	-18.00	-6.00	-4.67	1.19	7.00
20061213	24.00	3.00	2.07	0.43	3.00

The changes in the propagation latitude and longitude are shown in Figure 18 and Figure 19 respectively. The Coned model version 1.3 for these six extra cases were completed using the single-shot method utilized by Emmons [2012] where only a single value was used for each input which was the median of the 100 runs that were done for each case. This did not allow statistical analysis on these runs so standard deviations were not available for these cases.

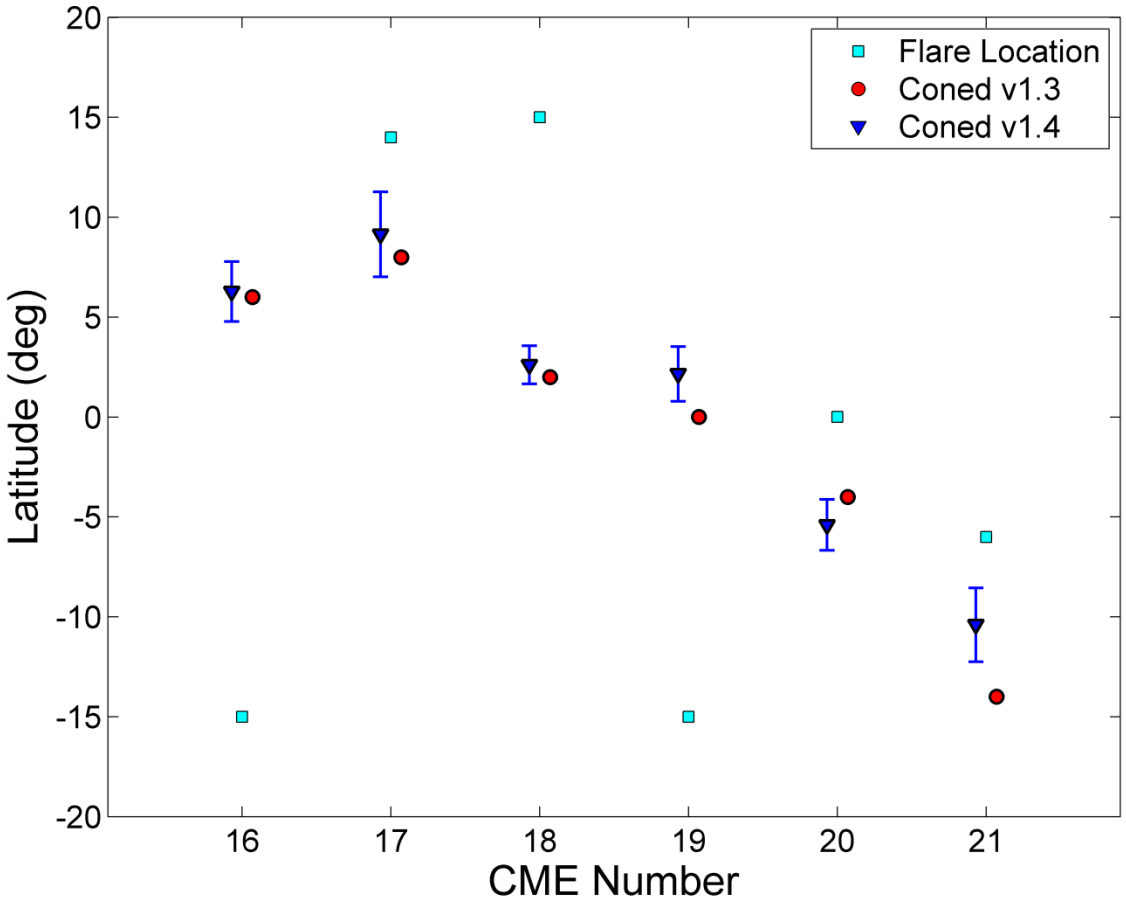

Figure 18: Comparison between the calculated cone latitude for the Coned model version 1.4 (with std error bars) and the original Coned model version 1.3 (using single-shot runs) with the flare location noted for reference for the six extra CMEs. The symbols are offset slightly to allow differentiation between values for the same CME.

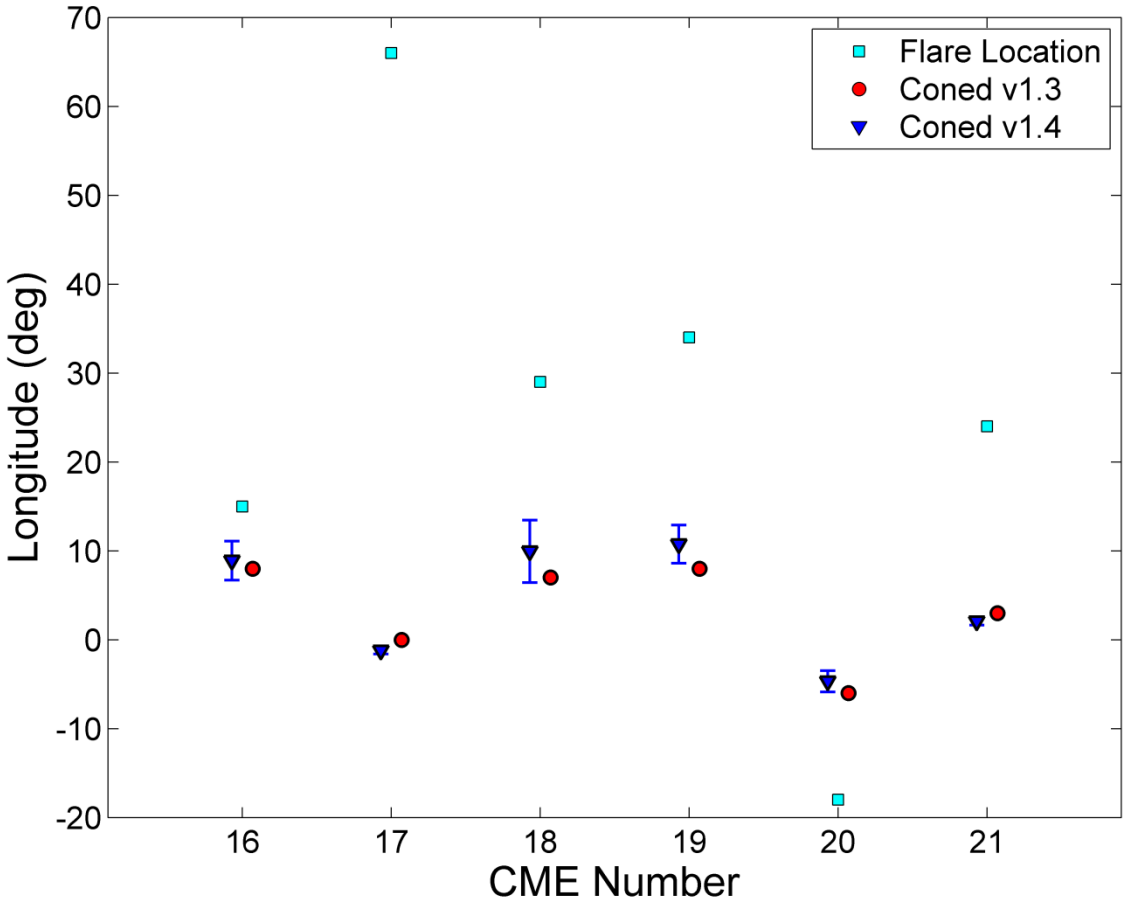

Figure 19: Comparison between the calculated cone longitude for the Coned model version 1.4 (with std error bars) and the original Coned model version 1.3 (using single-shot runs) with the flare location noted for reference for the six extra CMEs. The symbols are offset slightly to allow differentiation between values for the same CME.

Propagation Time

The improvements in the Coned model greatly increased the accuracy of the forecast even in these six extra cases analyzed. The statistics of the results are displayed in Table 13 while the forecast errors are given in Table 14. The absolute average error was reduced by 25% over the single shot median analysis of the original Coned model version 1.3 down to 4.43 hours. Surprisingly, the absolute average error of the median displayed a much larger improvement of 40% from 5.91 hours down to only 3.58 hours.

Since the single-shot method utilizes the median results from the Coned model as inputs into ENLIL, the median results in this case might be a better comparison between the two methods. This result also displays that, at least with the improvements, the median and mean values of the propagation times are not interchangeable as they appeared to be in the previous version [Emmons, 2012].

Table 13: The statistics for the propagation time for the extra 6 CMEs using the single shot Coned model version 1.3 and the Coned model version 1.4. In this table, *avg* stands for average, *std* stands for standard deviation, *med* stands for median, and *mad* stands for median absolute deviation.

CME date (YYYYMMDD)	Actual (hours)	Coned v1.3 (hours)	Coned v1.4 Avg (hours)	Coned v1.4 STD (hours)	Coned v1.4 Median (hours)	Coned v1.4 MAD (hours)	Coned v1.4 Range (hours)
19980502	36.38	45.95	44.37	5.39	43.18	3.88	22.68
20000809	53.25	55.72	57.69	5.92	56.10	4.90	26.83
20011019	47.40	48.70	54.47	7.92	53.12	5.22	39.47
20011122	30.15	38.93	33.39	4.74	32.32	3.34	17.53
20031118	46.83	42.18	47.12	5.44	46.33	3.48	32.85
20061213	35.03	43.73	38.57	2.25	38.45	1.43	13.08

Table 14: The forecast errors and performance metrics for the propagation time of the extra 6 CMEs using the single-shot Coned model version 1.3 and the Coned model version 1.4. A negative value represents the predicted arrival time was earlier than the actual time.

CME date (YYYYMMDD)	avg-actual Coned v1.3 (hours)	avg-actual Coned v1.4 (hours)	actual inside avg ±1 std?	med-actual Coned v1.4 (hours)	actual inside med ±1 mad?	actual inside range?
19980502	9.57	7.99	no	6.80	no	yes
20000809	2.47	4.44	yes	2.85	yes	yes
20011019	1.30	7.07	yes	5.72	no	yes
20011122	8.78	3.24	yes	2.17	yes	yes
20031118	-4.65	0.29	yes	-0.50	yes	yes
20061213	8.70	3.54	no	3.42	no	yes
absolute mean	5.91	4.43		3.58		

The improvements not only showed a similar improvement on these extra CMEs analyzed, but actually showed similar statistical results as well. All but two of the actual propagation times fell within one standard deviation of the average. Additionally, the actual results were within the range for all results. Only two of the six results were not better after the improvements, the 9 August 2000 and the 19 October 2001 CMEs. In both of these cases, the predicted propagation times were slower than originally predicted. A large CME occurred within 24 hours before each CME which could have cleared out the interplanetary medium and allowed for a faster arrival time than predicted by the ENLIL analysis of only a single CME at a time [Skoug et al., 2004].

Overall, the analysis of these six CMEs successfully accomplished the goal of confirming the improvements in the original 15 CMEs were due to an actual improvement in the program rather than a specific fitting for the original runs. The improvement for the averages ranged from 43% for the original CMEs to 25% for the six additional CMEs. The improvement for the medians ranged from 45% for the original CMEs to 39% for the six additional CMEs. The mean absolute forecast error of the median ensemble results for all runs was improved by over 43% with a mean propagation error of the median forecast of 4.59 hours.

The full results can be seen in Figure 20 for the averages and Figure 21 for the medians for the complete 21 CMEs. When looking at all 21 CMEs together with the improvements to the model, the correlation between the average predicted and actual propagation times was 0.85 with a p-value of 0.00. The correlation between the median predicted and actual propagation times was 0.86 with a p-value of 0.00. This indicates a

strong correlation between the predicted and actual propagation times of these 21 CMEs for both the average and median values.

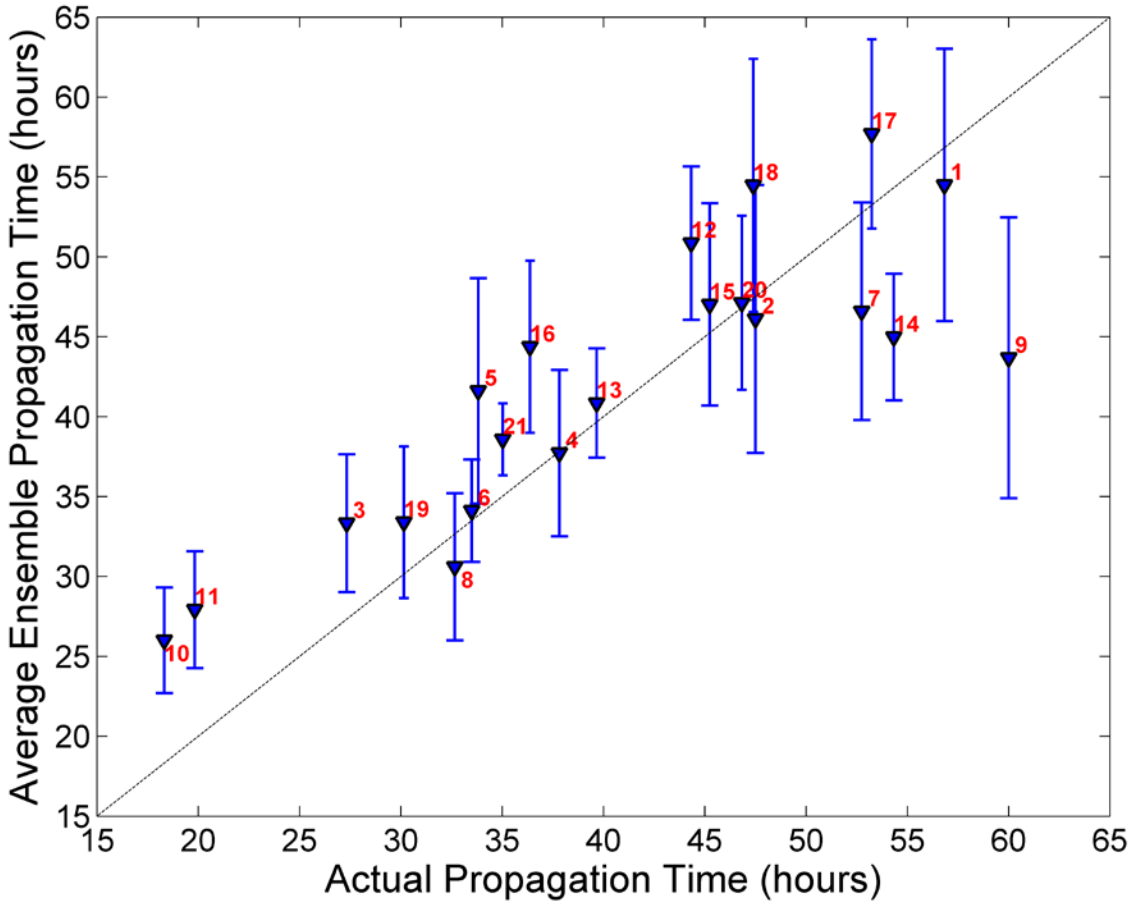

Figure 20: The averages and standard deviations of the ensemble propagation times versus the actual propagation times for the Coned model version 1.4 for all 21 CMEs.

When taken as a whole, there was a small but measurable improvement by using the median versus the mean for the ensemble runs. Time limitations put a cap on the amount of runs that can be reasonably done in a full ENLIL run where doubling the number of runs will double the run time. With only 100 runs and the random selection of pixels that the Coned model uses to determine the input parameters, there are

occasionally outliers in the data. These outliers affected the overall averages of the extra six CMEs more than the original 15 CMEs in this case as a few outliers changed the predicted propagation time by as much as 1.5 hours in one case as compared to the median. The effects of these outliers can be minimized by performing additional calculations or by using the median since the median is less sensitive to these outliers. Since prompt predictions are important in the types of real-world scenarios where these calculations would be necessary, it is recommended to use the median values for the final predictions in order to guarantee these outliers do not corrupt the predictions.

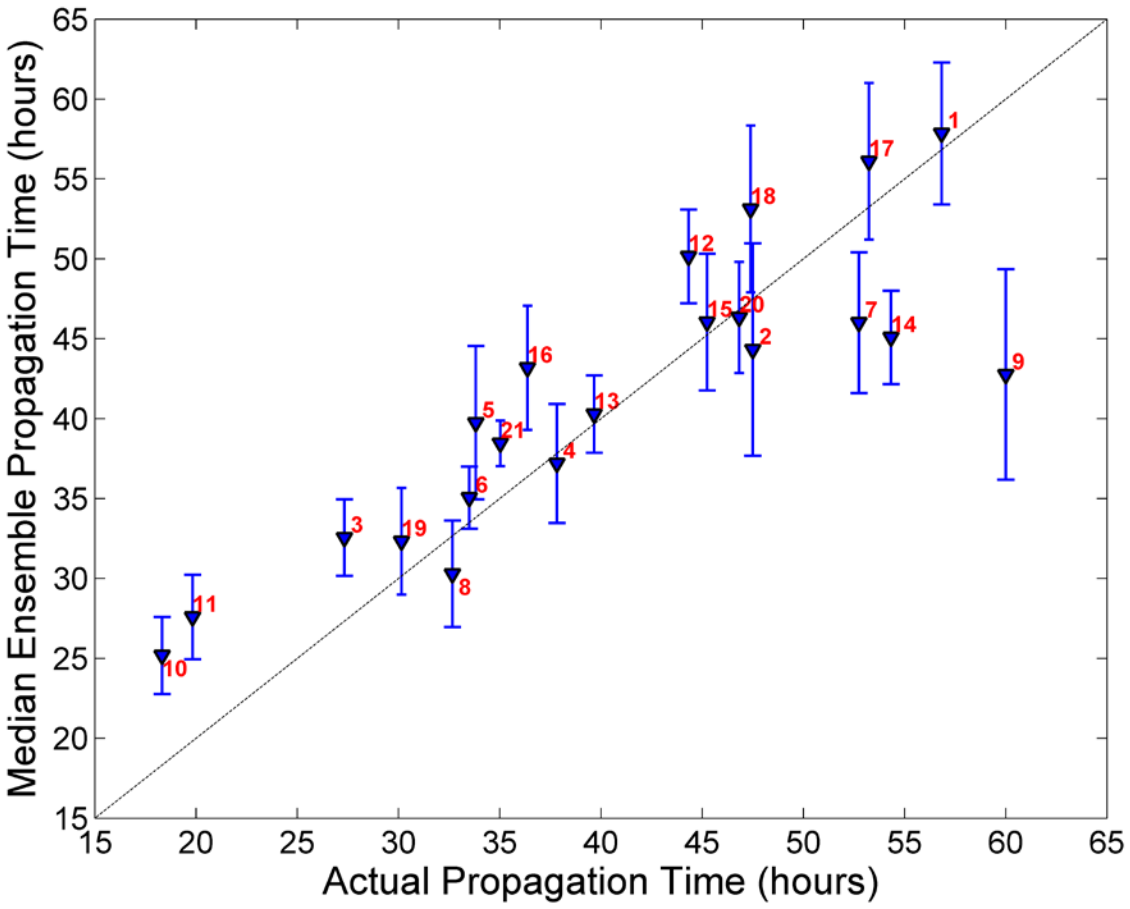

Figure 21: The medians and median absolute deviations of the ensemble propagation times versus the actual propagation times for the Coned model version 1.4 for all 21 CMEs.

Removal of the Climatological Weighting

The improvements to the Coned model version 1.3 decreased the errors in the predicted propagation times by around 40% on average using the median values but the climatological weighting in the program was still being used that was not based on observations of the CME. This weighting was added to push the opening half angle of the CME cone towards historical observations. In order to test if the improvements that were made to the Coned model provided enough information to be able to remove the climatological weighting from the code, single-shot runs were performed to see what the changes would be.

Input Parameters

The input parameters calculated from the Coned model version 1.4 before and after the climatological weightings are removed are given in Table 15. Overall, the removal of the weighting had little effect on the input parameters except in a single case, the 6 November 2004 CME. All other CMEs had the opening half angle change by three degrees or less. There were also no changes of more than one degree in any direction of the propagation axis of the CME cone except in the case previously mentioned. Finally, the changes in velocity were only greater than 5% in the two cases from 2003 and the 6 November 2004 CME.

Table 15: Comparison of the input parameters for the Coned model version 1.4 before and after the climatological weighting was removed. Here, *Lat-Lon* stands for calculated latitude-longitude of the propagation axis, *ω* stands for the calculated opening half angle of the CME cone given in degrees, *No CW* stands for no climatological weighting, and the change is color-coded with green representing less than 5% change, yellow is a 5-10% change, and red is a greater than 10% change.

CME date (YYYYMMDD)	Lat-Lon	Velocity	ω	Lat-Lon No CW	Velocity No CW	ω No CW	ΔVelocity	Δω-cloud
19990503	N12E27	839	84	N13E27	829	86	1.19%	2.38%
20000404	N01W17	1114	73	N01W17	1163	71	4.40%	2.74%
20000714	N03W06	1798	66	N03W05	1814	64	0.89%	3.03%
20010329	S00W00	1435	52	S00W00	1390	53	3.14%	1.92%
20010410	S12W02	1462	63	S12W02	1454	62	0.55%	1.59%
20010924	S08E22	1743	84	S07E21	1813	82	4.02%	2.38%
20011009	S09W03	1103	55	S10W03	1137	55	3.08%	0.00%
20011104	S00W07	1753	66	S00W07	1778	66	1.43%	0.00%
20011117	N03E16	1164	64	N03E16	1128	65	3.09%	1.56%
20031028	N00W00	2201	71	N00W00	2357	69	7.09%	2.82%
20031029	S04W02	2058	71	S03W02	2172	68	5.54%	4.23%
20040720	N00W00	1292	46	N00W00	1275	48	1.32%	4.35%
20041106	N04E02	1052	50	N06E02	1288	44	22.43%	12.00%
20041203	N08E03	1071	52	N09E03	1088	51	1.59%	1.92%
20100403	S03W02	965	44	S03W02	962	44	0.31%	0.00%
Average							**4.00%**	**2.73%**

Propagation Time

Comparable input parameters predictably produced similar predictions for the propagation times. The propagation time comparison for the Coned model version 1.4 before and after the climatological weighting was removed is given in Table 16. Overall, the results were very similar in both cases. Out of the 15 CMEs analyzed, there were only three cases where the predicted propagation times varied by more than one hour and there were no cases where the change in predicted propagation time was more than two hours. The absolute average error of the propagation times changed by only 0.21 hours upon removal of the climatological weight. These results imply that the climatological

weights for the opening half angles can safely be removed from further iterations of the Coned model.

Table 16: Comparison of the propagation times for the Coned model version 1.4 before and after the climatological weighting was removed. Here, *No CW* stands for no climatological weighting.

CME date (YYYYMMDD)	Actual (hours)	Median (hours)	Median-Actual Difference (hours)	Median No CW (MNC) (hours)	MNC-Actual Difference (hours)
19990503	56.83	57.85	1.02	57.62	0.79
20000404	47.50	44.32	-3.18	45.75	-1.75
20000714	27.33	32.55	5.22	33.12	5.79
20010329	37.83	37.19	-0.64	37.23	-0.60
20010410	33.83	39.74	5.91	40.20	6.37
20010924	33.50	35.05	1.55	34.13	0.63
20011009	52.75	46.00	-6.75	44.32	-8.43
20011104	32.67	30.28	-2.39	28.88	-3.79
20011117	60.00	42.76	-17.24	42.63	-17.37
20031028	18.33	25.18	6.85	25.30	6.97
20031029	19.83	27.59	7.76	27.60	7.77
20040720	44.33	50.15	5.82	49.83	5.50
20041106	39.67	40.28	0.61	40.70	1.03
20041203	54.33	45.08	-9.25	44.68	-9.65
20100403	45.25	46.04	0.79	46.92	1.67
Absolute Average			5.00		5.21

Effect of Multiple CMEs on WSA-ENLIL with Coned Model

In some cases, there were additional CMEs that occurred before or after the studied CME that might have an effect on its propagation time. In order to test this effect, the ENLIL code was run with multiple Coned model inputs from different CMEs at the same time to test the effects a previous CME could have on the following CME. For this run, the 29 October 2003 was used because the input parameters for the previous

large CME were already calculated, there were multiple CMEs within a short period of time, and it also was one of the worst predictions.

When the data was analyzed for this run, it was noticed that the density of the interplanetary medium was higher on the Earth side of the Sun. This can be seen in one of the ENLIL output density graphs shown in Figure 22. Additionally, the density behind the CME front is much lower than the density of the interplanetary medium that has not yet been hit by the CME.

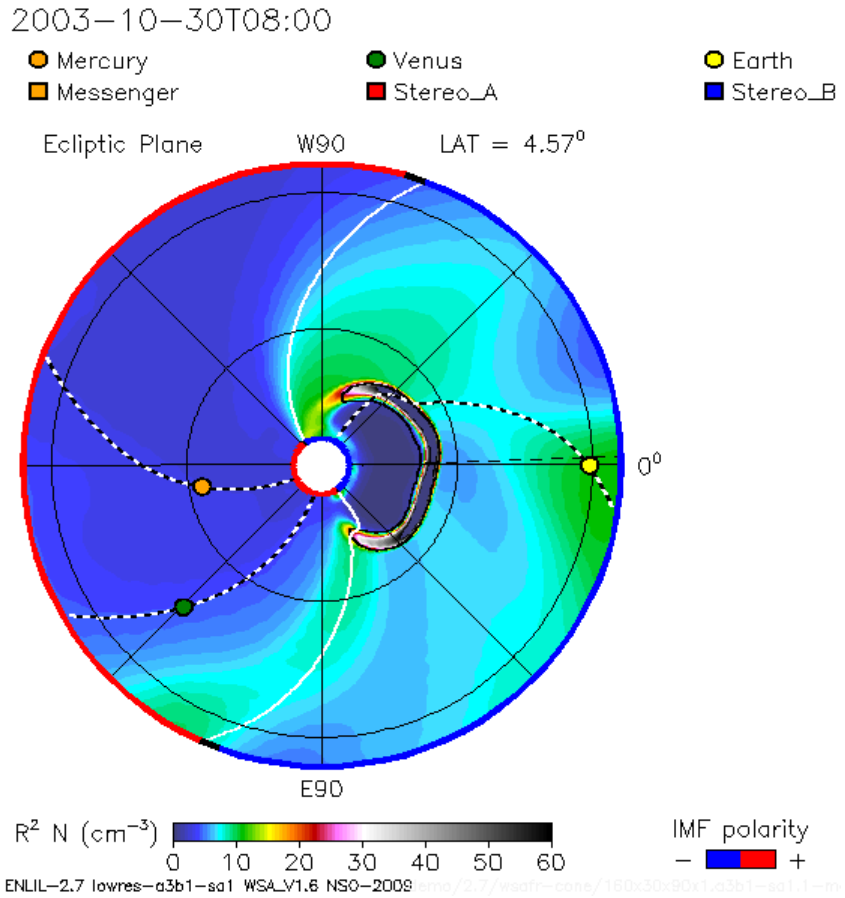

Figure 22: WSA-ENLIL output density graph for the 29 October 2003 CME shortly after the CME erupted.

WSA-ENLIL was run again incorporating both the 28 October 2003 and 29 October 2003 CMEs together in a single ENLIL run with both CMEs used as input. Two images from the results are displayed in Figure 23. The first image shows the reduced density behind the first CME that would normally slow the CME down during its transit from the Sun to the Earth. The second image shows the 29 October 2003 CME during the same time as Figure 22. Here, the density of the ambient solar wind in the path of the CME is several times lower than what it was before the additional CME was included in the ENLIL run. This prevents the second CME from decelerating as much as in the case where the CME was run individually and the 29 October CME arrives at Earth faster.

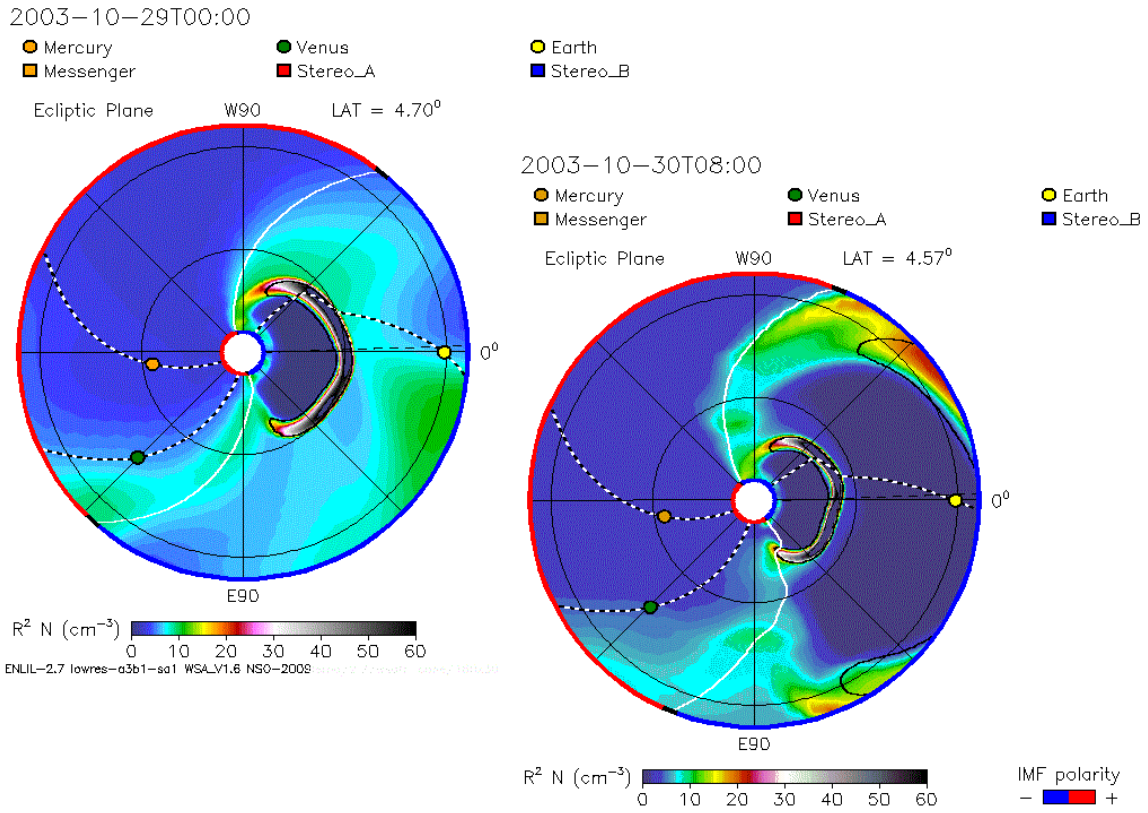

Figure 23: WSA-ENLIL output density graph for the 28 October 2003 CME (left) and 29 October 2003 CME (right) shortly after each CME erupted when ENLIL is run with both together. The edge of the 28 October 2003 CME can be seen at the edge of the figure on the right.

The arrival time of the 29 October 2003 CME was approximated as 1830 on 30 October 2003. The total predicted propagation time of this CME was 21.60 hours compared to the actual propagation time of 19.83 hours. This reduced the error of the predicted propagation time from 8.09 hours to only 1.77 hours. Combining the CMEs together in the same ENLIL run successfully moved the error on this CME from the third worst prediction of propagation time out of all 21 CMEs analyzed to one of the best.

Previous CMEs have a large and lasting impact on the background and therefore the propagation times on future CMEs in ENLIL. Adding additional CME inputs into ENLIL appears to provide a much more accurate representation of the propagation of the CME and seems to give a more accurate prediction of the impact of the CME on Earth as well than running the CMEs individually.

V. Conclusions

The mean absolute forecast error of the average for the 15 CMEs was improved by 43% from 9.06 hours to 5.16 hours. The mean absolute forecast error of the median for the 15 CMEs was improved by 45% from 9.17 to 5.00 hours. The ensemble forecast of the core CMEs predicted the propagation times of 8 out 15 events with enough accuracy that the propagation time fell within the ensemble average plus or minus the ensemble standard deviation. The original work by Emmons [2012] using the Coned model version 1.3 found only 5 out of the 15 events within one standard deviation of the ensemble average. Additionally, all five of those CMEs took between 30 and 46 hours to reach the Earth. With the improvements to the Coned model, the CMEs with propagation times as high as 57 hours were predicted accurately. The number of CMEs whose actual propagation time fell within the range of the ensemble also increased from 8 to 13.

For the complete set of 21 ensemble runs, the results proved similar. The actual propagation time fell within one standard deviation of the predicted value for 12 of 21 CMEs tested and 19 of 21 had the actual propagation time fall somewhere within the range of values predicted by the ensemble. Additionally, 18 of 21 CMEs showed an improvement in the accuracy of the prediction of the propagation time and 20 of 21 CMEs had a mean absolute error of the propagation time of less than 10 hours. For the full 21 CMEs, the mean absolute error of the average predicted propagation time was 4.95 hours. The mean absolute error of the median predicted propagation time was 4.59 hours. The median values provided a better prediction in most cases and mitigated the occasional outliers that occur in taking a random sample of pixels from the LASCO C3

difference images. This problem with outliers is corrected with larger sample sizes or by using the median values rather than the mean.

Additionally, the Coned model version 1.4 was made to be as automated and simple as possible. The only data needed were the LASCO C3 difference images and the location of the associated flare input into the program. The previous work with the Coned model version 1.3 was made by manually tweaking to the filtering threshold on each CME [Emmons, 2012]. The need for this has been removed and the process is now designed to minimize user bias.

The worst result, 17 November 2001, was the only CME with a forecast error greater than 10 hours. This CME had an error in predicted propagation time of 16.32 hours and was predicted to arrive early. This particular run might represent the limitations present in the Cone model. Since the Cone model approximates the CME as a cone, when the CME has a shape wildly different from a cone, as in this case, the input parameters cannot be calculated accurately.

Next, the climatological weighting was removed from the Coned model. The removal of this weighting for the opening half angle of the CME cone did cause a slight decrease in the accuracy of the predicted propagation times, but the mean absolute error of the forecasted median propagation time only increased by about 4% from 5.00 hours to 5.21 hours. The removal of this weighting appears to be possible now with only a minor impact in forecasting accuracy.

There were also a few cases where the predicted mean and median times for the arrival of the CME were six or more hours later than it actually arrived. These were the 2 May 1998, 10 April 2001, 28 October 2003, and 29 October 2003. All of these had

additional CMEs occurring prior to them. The 29 October 2003 event was tested with multiple CMEs and the forecasting error was reduced from 8.09 hours to only 1.77 hours. Including multiple CMEs in the same ENLIL run appears to reduce the forecasted transit time of the CME and would improve all of these runs.

Future Efforts

The next steps in ensemble forecasting of CMEs using WSA-ENLIL with the Coned model should be to continue with the investigation into the calculations of the propagation times of the CMEs with multiple CMEs occurring over a short period of time. The preliminary results were quite promising with the forecasting error of the 29 October 2003 CME reduced to under two hours. This would allow for the changes created by the CME mass passing through the interplanetary medium to be incorporated into the later forecast and improving the prediction for the propagation times.

Additionally, the Coned model is currently unable to recognize multiple CMEs in a single LASCO C3 difference image. This represents a limitation in the current iteration of the Coned model and limits some cases where a CME occurs very soon after another CME and is still present on the LASCO C3 difference images. Including the ability to recognize and separate these multiple CMEs on the same image would allow predictions of the propagation times in these cases as well.

Bibliography

Afraimovich, E., V. Demyanov, and T. Kondakova, Degradation of GPS Performance in Geomagnetically Disturbed Conditions, *GPS Solutions*, 7 (2), 109-119, 2003.

Arge, C., and V. Pizzo, Improvement in the prediction of solar wind conditions using near real time solar magnetic field updates, *Journal of Geophysical Research, 105*, 10, 2000.

Boteler, D., R. Pirjola, and H. Nevanlinna, The effects of geomagnetic disturbances on electrical systems at the Earth's surface, *Advances in Space Research, 22* (1), 17-27, 1998.

British Petroleum (BP) Statistical Review of World Energy June 2012. http://www.bp.com/assets/bp_internet/globalbp/globalbp_uk_english/reports_and_publications/statistical_energy_review_2011/STAGING/local_assets/pdf/statistical_review_of_world_energy_full_report_2012.pdf Accessed 5 Dec 2012.

Chen, P., Coronal mass ejections: Models and their observational basis, *Living Reviews in Solar Physics*, 8, 2011.

Cyr, O., et al., Properties of coronal mass ejections: Soho lasco observations from January 1996 to June 1998, *Journal of Geophysical Research, 105* (A8), 18, 169-18, 2000.

Emmons, Daniel J.,*Ensemble forecasting of coronal mass ejections using the WSA-Enlil with Coned Model*. MS thesis, AFIT/APPLPHY/ENP/12-M04. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2012.

Emslie, A., et al., Energy partition in two solar flare/CME events, *Journal of Geophysical Research, 109* (A10), A10, 104, 2004.

Falkenberg, T., A. Taktakishvili, A. Pulkkinen, S. Vennerstrom, D. Odstrcil, D. Brain, G. Delory, and D. Mitchell, Evaluating predictions of ICME arrival at earth and mars, *Space Weather, 9* (9), S00E12, 2011.

Forbes, T., A review on the genesis of coronal mass ejections, *Journal of Geophysical Research*, 105 (23153), 79, 2000.

Forbes, T., et al., CME Theory and Models, *Coronal Mass Ejections*, pp. 251-302, 2006.

Gopalswamy, N., A. Lara, R. Lepping, M. Kaiser, D. Berdichevsky, and O. Cyr, Interplanetary acceleration of coronal mass ejections, *Geophysical Research Letters, 27* (2), 145-148, 2000.

Gopalswamy, N., A. Lara, S. Yashiro, S. Nunes, and R. Howard, Coronal mass ejection activity during solar cycle 23, in *Solar Variability as an Input to the Earth's Environment*, vol. 535, pp. 403-414, 2003.

Illing, R., and A. Hundhausen, Observation of a Coronal Transient from 1.2 to 6 Solar Radii, *Journal of Geophysical Research, 90* (A1), 275-282, 1985.

Jackson, B., Imaging of coronal mass ejections by the Helios spacecraft, *Solar physics, 100* (1), 563-574, 1985.

Klein, L., and L. Burlaga, Interplanetary Magnetic Clouds at 1 AU, *Journal of Geophysical Research, 87* (A2), 613-624, 1982.

Newell, P., T. Sotirelis, K. Liou, C. Meng, and F. Rich, A Nearly Universal Solar Wind-Magnetosphere Coupling Function Inferred from 10 Magnetospheric State Variables, *Journal of Geophysical Research, 112*, 1-16, 2007.

Odstrcil, D., and V. Pizzo, Three-Dimensional Propagation of Coronal Mass Ejections (CMEs) in a Structured Solar Wind Flow: 1. CME Launched Within the Streamer Belt, *Journal of Geophysical Research, 104* (A1), 483-492, 1999.

Odstrcil, D., Modeling 3-D Solar Wind Structure, *Advances in Space Research, 32* (4), 497-506, 2003.

Odstrcil, D., P. Riley, and X. Zhao, Numerical Simulation of the 12 May 1997 Interplanetary CME Event, *Journal of Geophysical Research*, 109, A02, 116, 2004.

Owens, M., P. Cargill, C. Pagel, G. Siscoe, and N. Crooker, Characteristic Magnetic Field and Speed Properties of Interplanetary Coronal Mass Ejections and Their Sheath Regions, *Journal of Geophysical Research, 110*, 105, 2005.

Prölss, G., *Physics of the Earth's Space Environment: An Introduction*, Springer Verlag, 2004.

Pulkkinen, A., T. Oates, and A. Taktakishvili, Automatic Determination of the Conic Coronal Mass Ejection Model Parameters, *Solar Physics, 261* (1), 115-126, 2010.

Pulkkinen, A., private communication, 2012.

Rycroft, M. J., and S. K. Runcorn (Eds.), *The Solar Corona*, 1973.

Skoug, R. M., et al. "Extremely high speed solar wind: 29–30 October 2003." Journal of Geophysical Research 109.10.1029 (2004).

St. Cyr, O.C., Burkepile, J. T., Hundhausen, A. J. and Lecinski, A. R., 1999, A Comparison of Ground-Based and Spacecraft Observations of Coronal Mass Ejections from 1980–1989, *Journal of Geophysical Research*, 104 (A6), 12,493–12,506.

Taktakishvili, A., A. Pulkkinen, P. MacNeice, M. Kuznetsova, M. Hesse, and D. Odstrcil, Modeling of Coronal Mass Ejections That Caused Particularly Large Geomagnetic Storms Using ENLIL Heliosphere Cone Model, *Space Weather, 9* (6), S06,002, 2011.

Tascione, T. F., *Introduction to the space environment*, Krieger Publishing Company Malabar, Florida, 2010.

Tóth, G., and D. Odstrcil, Comparison of Some Flux Corrected Transport and Total Variation Diminishing Numerical Schemes for Hydrodynamic and Magnetohydrodynamic Problems, *Journal of Computational Physics, 128* (1), 82-100, 1996.

Vourlidas, A., Buzasi, D., Howard, R.A. and Esfandiari, E., 2002, "Mass and Energy Properties of LASCO CMEs", in *Solar Variability: From Core to Outer Frontiers*, Proceedings of the 10th European Solar Physics Meeting, Prague, Czech Republic, 9–14 September 2002, (Ed.) Wilson, A., vol. SP-506 of ESA Conference Proceedings, pp. 91–94, ESA Publications Division, Noordwijk.

Wang, Y., N. Sheeley, et al., Observations of Correlated White-Light and Extreme-Ultraviolet Jets from Polar Coronal Holes, *The Astrophysical Journal*, *508*, 899, 1998.

Wang, Y., and J. Zhang, A Comparative Study Between Eruptive X-class Flares Associated with Coronal Mass Ejections and Confined X-class Flares, *The Astrophysical Journal, 665*, 1428, 2007.

Webb, D., and A. Hundhausen, Activity Associated with the Solar Origin of Coronal Mass Ejections, *Solar Physics, 108* (2), 383-401, 1987.

Xie, H., L. Ofman, and G. Lawrence, Cone Model for Halo CMEs: Application to Space Weather Forecasting, *Journal of Geophysical Research, 109* (A3), A03, 109, 2004.

Yashiro, S., N. Gopalswamy, G. Michalek, O. Cyr, S. Plunkett, N. Rich, and R. Howard, A Catalog of White Light Coronal Mass Ejections Observed by the SOHO Spacecraft, *Journal of Geophysical Research, 109* (A7), A07, 105, 2004.

Zhang, J., K. Dere, R. Howard, M. Kundu, and S. White, On the temporal relationship between coronal mass ejections and flares, *The Astrophysical Journal*, *559*, 452, 2001.

Zhou, G., Wang, J. and Cao, Z., “Correlation between halo coronal mass ejections and solar surface activity”, *Astronomy and Astrophysics*, 397, 1057–1067, 2003.

Vita

Lt Jack Arthur Shepherd III was born outside of Chicago, Illinois on May 11, the son of Mr. and Mrs. Shepherd. After completing his work at Lexington High School in Lexington, Ohio, he went on to the Ohio State University where he double majored in Astronomy and Physics and received his Bachelor of Science in May 2004. He then travelled to Houston, Texas where he earned his Master of Arts in Teaching in January 2007 after which he became a high school math teacher. He entered the active duty Air Force in January 2009 and in August 2011 entered the Applied Physics Master's Degree program at the Air Force Institute of Technology at Dayton, Ohio.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
21-03-2013	Master's Thesis	Aug 2011 – Mar 2013

4. TITLE AND SUBTITLE
Optimization of Coronal Mass Ejection Ensemble Forecasting Using WSA-ENLIL with Coned Model

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Shepherd III, Jack A., 1st Lieutenant, USAF

5d. PROJECT NUMBER
N/A

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way, Building 640
WPAFB OH 45433

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENP-13-M-31

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

NASA/GSFC	AFWA
Dr. Antti Pulkkinen	Lt. Col. Robb Randall
NASA - Goddard Space Flight Center	101 Nelson Drive
Greenbelt, MD 20771	Offutt AFB, NE 68113-4020
301-286-0652 antti.a.pulkkinen@nasa.gov	271-9747 robb.randall@offutt.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA/GSFC AFWA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The combination of the Wang-Sheeley-Arge (WSA) coronal model, ENLIL version 2.7, and the Coned model version 1.4 was utilized to form ensemble forecast for 21 coronal mass ejections (CMEs). The Coned model was improved by adding a weight for the associated flare location to push the propagation axis towards the flare location and an additional image was used. Fifteen CMEs were used from previous research using the Coned model version 1.3 and the results were compared to determine the improvement. The model was used again with six additional CMEs and compared to the old Coned model version 1.3 to verify the results. The mean absolute forecast error of the median ensemble results from the original 15 CMEs was improved by 43% over the original Coned model version 1.3. The six additional CMEs verified the improvements. The model was used again to look at multiple CMEs at the same time to get a result where the multiple CMEs occurred over a short period of time. This resulted in an improvement in the error of a sample CME from 8.09 hours to 1.77 hours.

15. SUBJECT TERMS
Coronal Mass Ejection, Ensemble Forecasting, WSA-ENLIL with Coned Model

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Acebal, Ariel O., Lt Col, USAF
U	U	U	U	87	19b. TELEPHONE NUMBER *(Include area code)* (937) 255-3636, x 4518 (ariel.acebal@afit.edu)